①

Technical Report 657

Field Application of Four Models Designed to Predict Transfer-of-Training Potential of Training Devices

Dennis G. Faust, Robert W. Swezey, and Kenneth W. Unger
Behavioral Sciences Research Center
Science Applications, Inc.

Training and Simulation Technical Area
Training Research Laboratory

U. S. Army
Research Institute for the Behavioral and Social Sciences

October 1984

DISCLAIMER NOTICE

U. S. ARMY RESEARCH INSTITUTE

FOR THE BEHAVIORAL AND SOCIAL SCIENCES

A Field Operating Agency under the Jurisdiction of the

Deputy Chief of Staff for Personnel

L. NEALE COSBY
Colonel, IN
Commander

EDGAR M. JOHNSON
Technical Director

Research accomplished under contract
for the Department of the Army

Science Applications, Inc.

Technical review by

Warren Simmons
Robert Hays

NOTICES

<u>DISTRIBUTION:</u> Primary distribution of this report has been made by ARI. Please address correspondence concerning distribution of reports to: U.S. Army Research Institute for the Behavioral and Social Sciences, ATTN: PERI-TST, 5001 Eisenhower Avenue, Alexandria, Virginia 22333.

<u>FINAL DISPOSITION:</u> This report may be destroyed when it is no longer needed. Please do not return it to the U.S. Army Research Institute for the Behavioral and Social Sciences.

<u>NOTE:</u> The findings in this report are not to be construed as an official Department of the Army position, unless so designated by other authorized documents.

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER ARI Technical Report 657	2. GOVT ACCESSION NO. AD-A170 544	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) FIELD APPLICATION OF FOUR MODELS DESIGNED TO PREDICT TRANSFER-OF-TRAINING POTENTIAL OF TRAINING DEVICES		5. TYPE OF REPORT & PERIOD COVERED Final Report April 1983-February 1984
		6. PERFORMING ORG. REPORT NUMBER SAI-84-06-178
7. AUTHOR(s) Dennis G. Faust, Robert W. Swezey, Kenneth W. Unger		8. CONTRACT OR GRANT NUMBER(s) MDA 903-83-C-0177
9. PERFORMING ORGANIZATION NAME AND ADDRESS Behavioral Sciences Research Center Science Applications, Inc. McLean, VA 22102		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS 2Q263744A795
11. CONTROLLING OFFICE NAME AND ADDRESS U.S. Army Research Institute for the Behavioral and Social Sciences 5001 Eisenhower Avenue, Alexandria, VA 22333-5600		12. REPORT DATE October 1984
		13. NUMBER OF PAGES 90
14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office) --		15. SECURITY CLASS. (of this report) Unclassified
		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE --

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

--

18. SUPPLEMENTARY NOTES

Contracting Officer's Representative, Angelo Mirabella

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Transfer of training	Training model
TRAINVICE	Training evaluation
Simulation	Training effectiveness
Analytical model	Forecast model

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

This report describes the field application of four analytical models designed to predict transfer of training from simulators to the operational equipment for which the simulators were designed. The four analytical models were applied to two prototype Army Maintenance Training Simulation System (AMTESS) simulators. Predictions of transfer of training derived from application of the models were compared to transfer-of-training data obtained in an experimental evaluation of the two prototype simulators. Discussion and comment

(Continued)

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

UNCLASSIFIED

 SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

A

Technical Report 657

Field Application of Four Models Designed to Predict Transfer-of-Training Potential of Training Devices

Dennis G. Faust, Robert W. Swezey, and Kenneth W. Unger
Behavioral Sciences Research Center
Science Applications, Inc.

Contracting Officer's Representative
Angelo Mirabella

Submitted by
Robert J. Seidel, Chief
Training and Simulation Technical Area

Approved as technically adequate
and submitted for publication by
Harold F. O'Neil, Jr., Director
Training Research Laboratory

U.S. ARMY RESEARCH INSTITUTE FOR THE BEHAVIORAL AND SOCIAL SCIENCES
5001 Eisenhower Avenue, Alexandria, Virginia 22333-5600

Office, Deputy Chief of Staff for Personnel
Department of the Army

October 1984

Army Project Number
2Q263744A795

Training and Simulation

ARI Research Reports and Technical Reports are intended for sponsors of R&D tasks and for other research and military agencies. Any findings ready for implementation at the time of publication are presented in the last part of the Brief. Upon completion of a major phase of the task, formal recommendations for official action normally are conveyed to appropriate military agencies by briefing or Disposition Form.

FOREWORD

This research compared four models designed to predict the transfer-of-training potential of two forms of training devices and three categories of maintenance MOS. Because the study was limited by a number of conditions imposed by the field setting, generalizability is limited and the findings should be regarded as preliminary estimates of the validity and reliability of the models. This study does, however, provide future researchers with hypotheses regarding the metric properties and practical utility of the models and describes possible problems that might be encountered by those who attempt to replicate the findings in a comparable setting. Also, recommendations are given for a more controlled assessment of the models.

Edgar M. Johnson

EDGAR M. JOHNSON
Technical Director

ACKNOWLEDGMENTS

In addition to those acknowledged on the cover, the completion of this study was made possible through assistance and input provided by many individuals. For their contributions, acknowledgment is given to Martha J. Cole and M. Joseph Moyer of SAI, who served as field researchers in the study effort. Ms. Cole provided additional helpful information concerning insights about field data collection and a critique of earlier versions of the report. The authors also wish to thank Robert T. Hays and Angelo Mirabella of the U.S. Army Research Institute.

FIELD APPLICATION OF FOUR MODELS DESIGNED TO PREDICT TRANSFER-OF-TRAINING POTENTIAL OF TRAINING DEVICES

EXECUTIVE SUMMARY

Requirement:

This effort applied four models designed to predict the transfer-of-training potential of training devices to two prototype generic maintenance simulators in order to assess the measurement properties of the models, and to obtain feedback on the practical utility of each model.

Procedures:

Two prototype Army Maintenance Training and Evaluation Simulation System (AMTESS) simulators served as the test bed. Students in three similar Military Occupational Specialties were trained on one of the simulators. Performance measures (transfer-of-training scores) were obtained on the students after completing simulator training, and used as criterion measures. Each of the four transfer-of-training models was then applied to the simulators to produce a transfer-of-training prediction. The predictions were then correlated with the criterion measures, and reliability estimates were obtained on each model. Analysts who applied the models were surveyed to secure feedback as to the practical utility of the models.

Findings:

The summary predictions produced by each model to estimate the transfer-of-training potential of a simulator proved to be misleading. However, when predictions were made for each independent task and these task-level predictions were independently correlated with the criterion, prediction improved. Nevertheless, the predictive power of each model was weak--the two most predictive models correlating only .33 and .34 with the criterion. Surprisingly, the reliability of each model was high, indicating potential measurement contamination. Further, analysts who applied the models reported that they were complex, difficult, and time-consuming to use, and should be simplified.

Utilization of Findings:

The results of this effort were severely confounded by conditions imposed by the field setting. Generalizability is therefore limited. Findings should be regarded as tentative, but can serve as hypotheses for future research regarding each model's validity. Recommendations are given for a more controlled assessment of the four models.

FIELD APPLICATION OF FOUR MODELS DESIGNED TO PREDICT TRANSFER-OF-TRAINING POTENTIAL OF TRAINING DEVICES

CONTENTS

LIST OF TABLES

LIST OF FIGURES

I. INTRODUCTION

BACKGROUND

Basic to the development of new military systems is the training of personnel to man those systems; including the decision whether to "device-train" (in whole or part), and what training device offers the best benefit to the Army. The ability to prescribe and/or evaluate training devices in this regard has been a subject of Army study for several years and for good reason. The greater complexity, costs, and hazards associated with use of actual field equipment for training has increased reliance upon training devices. Further, the trend toward training device use has been encouraged by the advent of new technologies such as developments in micro-circuitry, micro-processors, training hardware and software, etc. Thus, training devices and simulators offer various training benefits and cost savings to the military.

In the 1970's, the U.S. Army Research Institute (ARI) initiated work to develop a methodology for evaluating the transfer-of-training potential of training devices. A preliminary model for accomplishing this end was produced by Wheaton, Fingerman, Rose and Leonard (1976a) and Wheaton, Rose, Fingerman, Korotkin and Holding (1976b) and became known as TRAINVICE. The model purported to provide a "...feasible and reliable set of procedures for processing the data to generate predictions of potential training device effectiveness" (Wheaton et al., 1976a, p. 3). Further efforts were sponsored by ARI to advance the methodology for either predicting device efficacy or for prescribing device requirements. Beyond the original model, three additional models evolved over the years (i.e., Hirshfeld and Kochevar, 1979; Narva, 1979a,b; Swezey and Evans, 1970).[1] All four models shared certain common elements (at least in concept). Still, each differed in various ways from the others, (e.g., in required inputs, metrics, computational procedures). Although all of the models made some contribution to predicting transfer of training, it became clear that much work remained to be done to produce a model which was valid and truly state-of-the-art. The study reported in this document contributes to those efforts by providing findings on a field application of the four transfer-of-training models. Before addressing the present study, however, a brief description of each of the models will help to illustrate what has been accomplished to date, what remains to be accomplished, and the relevance of this present research.

The Original TRAINVICE Model (Wheaton et al., 1976a,b)

The original model (Wheaton et al., 1976a,b) purported to predict transfer-of-training potential (i.e., from device training to the field equipment). In the model, device efficacy is seen to be a function of three factors. The first of these factors concerns the trainee learning deficit

[1] Various literature (e.g., see Tufano and Evans, 1982) have referred to the four models as TRAINVICE models number 1, 2, 3, and 4, or as A, B, C, and D, and a fair amount of confusion has resulted. For this reason, we here cite the models by standard author citations.

to be overcome. The second concerns the training techniques employed by the device to overcome the learning deficit. The third concerns transfer potential of the learning as regards the operator subtasks trained. With respect to this last factor, the model espouses the theory of "identical elements" (Thorndike, 1903; Thorndike and Woodworth, 1901) and assesses the physical and functional similarity between the device characteristics and parent (field) equipment characteristics to make a transfer prediction. Additionally, it considers subtask overlap (communality) between the training device and parent equipment. The various factors above are assessed through the following five analyses:

1. Task communality analysis (C)

2. Physical similarity analysis (PS)

3. Functional similarity analysis (FS)

4. Learning deficit analysis (D)

5. Training techniques analysis (T)

For each of these analyses, an assessor assigns ratings on a subtask-by-subtask basis as per the operational tasking to be trained. The data developed are judgmental (from rating scale criteria) and the overall analysis requires considerable expertise and time to implement (e.g., an instructional psychologist might require weeks to complete a protocol for a training system of only moderate complexity).

Once all input data are amassed, a transfer-potential index is calculated for the device. The resulting index score resides in the range 0 to +1, with higher scores indicating a greater transfer potential. The mathematical model for the Wheaton et al. model is:

$$\text{Index} = \frac{\sum_{i=1}^{N} C_i \times S_i \times D_i \times T_i}{\sum_{i=1}^{N} D_i}$$

where:

C_i = task communality value

S_i = average of the physical and functional similarity assessments

D_i = learning deficit analysis value

T_i = training techniques analysis value

N = number of subtasks required in the training

The index does penalize a training device when it fails to cover a subtask required in the operational setting. However, it does not penalize the device if superfluous instruction is provided.

Hirshfeld and Kochevar, 1979

The Hirshfeld and Kochevar model is generally similar in concept to the original Wheaton et al. version. In the Hirshfeld and Kochevar version, however, device efficacy is seen as a function of essentially two factors: device characteristics and personnel training requirements. The model assesses these two factors through five analyses:

1. Task commonality analysis (TC)

2. Physical similarity analysis (PS)

3. Functional similarity analysis (FS)

4. Skill and knowledge requirements (SKR)

5. Task training difficulty (TTD)

Though the terms used in the analyses may appear similar to those of the Wheaton et al. model, specific differences exist. For example, this version assesses training requirements at the task level rather than at the subtask level. Further, it considers task elements which do and do not require device training, therefore penalizing the device when essential tasking is omitted or when superfluous tasking is included. The model is somewhat less time consuming to administer than the Wheaton, et al., version but not greatly so.

In Hirshfeld and Kochevar's model, the physical and functional similarity analyses are similar to those of the Wheaton et al. model. Also, the skill and knowledge requirements analysis generally corresponds to the learning deficit analysis of the Wheaton et al. model. However, the task training difficulty analysis is considerably different from the Wheaton et al. model in that it employs "training time" estimates as input data. Perhaps the most notable difference between the two models is that Hirshfeld and Kochevar's version makes no attempt to assess the training techniques employed by a device to effect instruction.

As with Wheaton et al., this version uses assessor judgment (via rating scale criteria) to quantify the necessary input data for the overall analysis. Here too, the resulting index score ranges from 0 to +1 with a higher score indicating greater transfer-of-training potential. Once the data are produced, the prediction is calculated through the following mathematical model:

$$\text{Index} = \frac{\sum_{i=1}^{N} \left(\frac{TC + PS + FS}{3}\right) \times \left(\frac{SKR + TTD}{2}\right)}{\sum_{i=1}^{N} \left(\frac{SKR + TTD}{2}\right)}$$

where:

TC = task commonality analysis value

PC = physical similarity analysis value

FS = functional similarity analysis value

SKR = skill/knowledge requirements value

TTD = task training difficulty

N = number of training tasks involved

The model does provide a correction factor which can be applied to the index to adjust it for any device tasking not required in the operational setting. The model thus can penalize a device for failing to cover essential tasking or for providing superfluous training.

Narva, 1979a,b

This third model provides a considerable modification to the original one. Here, the spirit of the Wheaton et al. model is retained; however, there is a definite shift in focus toward predicting device effectiveness (i.e., the ability of the device to fulfill training objectives under the assumption that effectiveness is a valid predictor of tranfer potential). Narva's model is built around three major input factors. The first concerns what must be trained on the device. The second factor essentially concerns the importance (Narva called this the "why") of the training to be covered, i.e., the proficiency the trainee must achieve and the corresponding difficulty in so doing. The third factor concerns how the device provides for instruction in the course of meeting the training objectives. This model obtains its input data through the following six analyses:

1. Coverage requirements analysis (CR)

2. Coverage analysis (C)

3. Training criticality analysis (C_i)
   (i.e., degree of proficiency required)

4. Training difficulty analysis (D)

5. Physical characteristics analysis (PC)

6. Functional characteristics analysis (FC)

Narva's model first identifies operator performance required on the actual field equipment. It then assesses whether the device "covers" those requirements and penalizes the device if it does not (no penalty is given if superfluous skills or knowledges are trained). Unlike earlier models which employed subtask or task element descriptions alone, this model defines coverage requirements in terms of skill and knowledge components subsumed within the tasks. The criticality (i.e., proficiency level which the trainee must achieve) for each skill/knowledge is then determined, along with the degree of difficulty required to learn each. Last, the physical and functional characteristics of device displays/controls are assessed for how well they support training. On the face, this latter analysis resembles the attempts of earlier models to consider identical elements between device and parent equipment (i.e., transfer potential). In fact, it is a training techniques analysis which essentially assumes the parent equipment to represent the optimal training medium. What this last analysis does assess is how well the stimulus and

response characteristics of the device support the training requirements in light of good instructional practices (i.e., of the ISD Learning Guidelines; see Aagard and Braby, 1976, and Branson, Raynor, Cox and Hannum, 1975). The model is equally as laborious and time consuming to administer as the original Wheaton, et al. model.

As with the earlier models, this version relies upon assessor judgment (rating scale criteria) to develop its input data for each of its component analyses. The index value (0 to +1) and its interpretation remain the same as for the prior two models. Once the data are developed, the mathematical model by which the forecast is computed is:

$$\text{Index} = \frac{\sum_{i=1}^{N} (CR \times C \times C_i \times D \times (PC + FC))_i}{\sum_{i=1}^{N} (CR \times C \times 4 \times 4 \times (PC_{max} + FC_{max}))_i}$$

where:

CR = skill/knowledge coveragements requirements

C = actual coverage provided by the device

C_i = training criticality value

D = training difficulty value

PC = physical characteristics analysis value

FC = functional characteristics analysis value

N = number of skills/knowledge required in the training

The demoninator values of "4" and the variables labeled "max" represent the ideal scores which the parent equipment would receive; the actual device scores being represented in the numerator. The final index (ranging from 0 to +1) is a proportion reflecting transfer-of-training potential by Narva's criterion, i.e., how well the device training approximates training accomplished on the parent equipment.

Swezey and Evans, 1980[2]

The Swezey and Evans model was originally commissioned as a user's guidebook to operationalize the propositions in Narva's (1979a,b) model. However, in the course of developing the guidebook, modifications to Narva's model were deemed necessary, and, therefore, were included (see Evans and Swezey, 1980). This, in fact, resulted in a substantially separate model. The concept behind the Swezey and Evans version is fundamentally the same as Narva's, the model being effected through the following analyses:

[2]The literature which originally described this model referred to it as "TRAINVICE II" because it represented the first effort to create a user's handbook and formal protocol for applying TRAINVICE methods.

1. Coverage analysis (C) (i.e., in light of coverage requirements)

2. Training proficiency analysis (P)

3. Learning difficulty analysis (D)

4. Physical characteristics analysis (PC)

5. Functional characteristics analysis (FC)

In the course of developing a guidebook for Narva's model, however, it became necessary to incorporate substantial modifications to Narva's rating scale definitions. In this version, input data for the analysis remain judgmental, as is the case with earlier models; however, more extensive scale definitions for classifying stimulus and response characteristics of device displays/controls were included, and the ISD learning guidelines were labeled to indicate whether they pertained to physical or functional device characteristics. Still, the model is equally time consuming to administer as the Narva model; requiring a great deal of specific input-data and considerable analyst knowledge of instructional principles.

Another major change incorporated by Swezey and Evans involved the mathematical algorithm for computing the transfer index. While the 0 to +1 index is employed as before and the model's components are generally similar to Narva's, the form used to compute the final index is:

$$\text{Index} = \frac{\sum_{i=1}^{N} \left(\frac{PC + FC}{PC_{max} + FC_{max}}\right) x\ (C \times P \times D)}{\sum_{i=1}^{N} (P \times D)}$$

where:

PC = physical characteristics analysis value

FC = functional characteristics analysis value

C = completeness of training coverage by the device

P = training proficiency student must attain

D = training difficulty value

N = number of skills/knowledges required in training

The closer the index approaches +1, the more the training provided by the device reflects that provided by the parent equipment. As with Narva's model, this version also penalizes requirements which a device fails to cover, but does not penalize when superfluous tasking is present. The design of the analyst's worksheet does, however, permit superfluous task coverage to be identified and listed.

RESEARCH AND DEVELOPMENT NEEDS

The descriptions of the four transfer-of-training models highlight the salient features of the existing methodology. What seems immediately notable, from even this brief review, are the various differences between the models. Although a conceptual "theme" appears to be present across their evolution, differences in theoretical constructs, component variables and mathematical formulations are apparent. Each model offers strengths and weaknesses, and it is difficult to conclude that all of the models could be equally effective in predicting transfer-of-training (or that one is superior) based on their face validity alone. Rather, what began as a development effort seems to have produced four different theoretical models with very little yet known concerning their respective validities.

Presently, the research and development needs of the models (to which this present report is intended to contribute) can be subsumed by essentially four problem areas needing resolution:

1. Theoretical construct of each model

2. Mathematical formulation (representing the relationships among construct variables)

3. Measurement issues (validity, reliability and precision)

4. Convenience of application (acceptability in practice)

An excellent in-depth review concerning the four models in relation to these R&D needs can be found in Tufano and Evans (1982). No attempt will be made to restate that work here. However, it is appropriate at this point to highlight some of these problem areas and other research accomplished to date in order to set the context for the present study.

Theoretical Construct. The four models do reflect a common set of assumptions in their framework. Essentially, this is:

1) Some learning deficit (subsuming the content and proficiency which training must achieve) must be overcome through...

2) training techniques which adequately deliver the training content to assure learning. Further...

3) similarity between device and parent equipment must be sufficient to permit transfer to occur.

These basic assumptions would appear to represent a fundamentally sound construct for predicting effective transfer. The problem is, ostensibly, that in developing this construct, each model has employed somewhat different variables and interpretations of those variables. Variability in this regard ranges from subtle to striking. It is difficult to imagine that these differences are inconsequential to prediction, and resolving these differ-

ences in favor of some truly valid construct is essential. Further, the problem of variability in construct definition is aggravated by the absence of a supporting data base for any of the models.

Mathematical Formulation. The four models each provide a mathematical formula intended to forecast transfer. Mathematically, each has specific weaknesses which seem to justify the need for improved modeling. To illustrate the mathematical problems involved in their formulations, consider one case - the area of "similarity" between device and parent equipment. This is a factor which all of the models seem to regard as fundamental to tranfer prediction. Though the specific models differ somewhat in how they calculate it, essentially their similarity (S_i) index is represented as a function (average) of two quantities: P, the physical similarity index, and F, the functional similarity index. Displays and controls receive a minimum score of 0 (not represented) and a maximum of some other positive integer (e.g., +3) when their fidelity approaches identity with the parent equipment. No provision is made in the mdoels for indicating misleading representation (i.e., producing negative scores); so that if a display or control is represented at all, no matter how badly or misleading, it obtains a positive similarity score. Since the mathematical integration of P and F varies between 0 and 1 in the models, S_i varies on that interval. With regard to the manner in which this similarity score is produced by the models' mathematics, two assumptions seem apparent:

- Physical and functional similarity are equally important to learning transfer.
- Any representation of a display or control, no matter how bad or misleading, can vary only on some positive value.

The first of these assumptions remains open to some question; the second would seem to provide the mathematical models with a blind spot of significance. As with the computation of the similarity index, the mathematical representations of "training techniques" effectiveness in at least three of the models also ignore the possibility of antagonistic device characteristics, proactive inhibition and negative transfer. These issues are but partially representative of a number of problems inherent in the mathematical formulations of the respective models.

Measurement Issues. In their present forms, each model can be regarded as a predictive selection device - a "test" of the training transfer potential of alternative training devices or device designs. Like all tests, the models are subject to the need for demonstrated validity, reliability, and precision. Based on no more than the brief overview of the four models, the prospective user is quickly prompted to ask which of the models possesses the higher degree of these metric properties. In-depth inspection of literature describing each model justifies even greater concern in this regard, due primarily to imprecise semantics of the models; i.e., questions of exactly what is being measured, what are the valid ranges of the variables, etc.

Since most of the initial research on the models necessarily focused on model design, very little study of metric properties has been conducted to date and that which has been conducted is essentially reducible to a

few studies (Wheaton, Rose, Finerman, Leonard and Boycan, 1976c; Rose, Wheaton, Leonard and Finerman, 1976; Swezey, Chitwood, Easley and Waite, 1977; and Swezey, 1983), all of which address only the original Wheaton et al. TRAINVICE model. Two other studies (Klein, Kane, Chinn and Jukes, 1978; Knerr, Nadler and Dowell, 1983) have examined applications of the Narva model and the Swezey and Evans model, respectively, but provide only subjective estimates on the validity of those models. A study by Faust (1982) examined the validity of the ISD Instructional Guidelines used to generate data for three of the models; however, that study addressed only the "training techniques analysis" component of those models. Thus, the classic measurement issues persist and no previous study has assessed all four models through any common test-bed to ascertain their metric properties or compare their precision.

Convenience of Application. In the final analysis, any valid measurement instrument must be convenient to implement in order to be of practical utility. Perhaps the least disputable problem concerning the four models is that all are impracticably difficult to apply to an existing training device or device design. Given a training device of only moderate complexity, for example, a complete analysis may require weeks (or even months) to complete; for a major system, perhaps a year. The device assessments required by each model are micro-analytic in nature and many in number. Developing an ADP-based computation system to run each model's calculations is a current area of interest intended to speed results and reduce administrative error of the models. Presently, however, no solution exists for the large volume of field data that must be collected to drive the models. Evaluation of device designs thus remains a labor intensive effort of impractical proportion. Applying the model in a "formative" mode (i.e., prescribing optimal device requirements from gradually evolving engineering specifications of a new weapon system) may be a more manageable application of the models. However, further research and development will be required before the models can be applied reliably so early in the systems development process.

STUDY PURPOSE

The TRAINVICE methodology consists of four models which share commonalities yet possess distinct differences. The four have progressed little beyong their embryonic stage and are labor intensive to apply. Relatively little field work has been done to determine their predictive properties and practical utility. Essentially, research needs reside in: 1) construct refinement, 2) mathematical modeling, 3) definition of the models' measurement properties, and 4) developing their convenience of application.

Current research initiatives are focusing on these areas to develop revised state-of-the-art models. Those developments should lead to field testing and perhaps the first extensive validity data on their application. Still, very little field-test data exist although such data would be useful to current R&D efforts. The opportunity to apply all four models to a common test-bed became available, however, as part of ARI's SIMTRAIN program of research with Science Applications, Inc. (SAI). (See for instance, Unger, Swezey, Hays, & Mirabella, 1984).

During the SIMTRAIN efforts, SAI was able to apply the four models to two breadboard maintenance simulators across four maintenance military

occupational specialities (MOSs). This study's purpose was to compare the four models in terms of their predictive efficacy and convenience of application. The first of these two objectives examined predictive differences between the four models (in relation to actual student transfer-of-training measures) and user reliability. The second, as stated, examined the practical convenience of applying the models. The four models were applied to the following two Army Maintenance Training and Evaluation Simulation System (AMTESS) maintenance training simulators:

- An AMTESS breadboard maintenance training device designed by a consortium of Seville Research Corporation and Burtek, Inc. (addressing training tasks involving a diesel engine)

- An AMTESS breadboard maintenance training device designed by the Grumman Aerospace Corporation (addressing tasks involving a self-propelled howitzer)

Copies of the two devices were located at both Aberdeen Proving Ground, Maryland, and at Fort Bliss, Texas. The four MOSs involved were:

- 63D30 - Self-propelled Field Artillery Systems Mechanic

- 63H30 - Direct Support Maintenance Supervisors

- 63W10 - Direct Support Vehicle Repairman

- 24C10 - Hawk Missile Firing Section Mechanic

Nine (9) contractor-trained analysts served as the subjects to apply the four transfer models to the MOSs and simulators.

Section II of this report describes the method employed to conduct the study. Section III reports study findings and Section IV discusses conclusions and recommendations. Where appropriate, relevant documentation is referenced and attached in the Appendix. Although findings of this study were subject to limitations imposed by the field setting and thus must be regarded as preliminary to full-scale validation of the models, the results hopefully will contribute to the transfer-of-training forecast initiative by providing basic insights regarding the predictive efficacy and practical utility of the respective models.

II. METHOD

OVERVIEW

The purpose of this study was to apply the four transfer-of-training models to each of two maintenance training devices; and to compare the models' predictions to actual transfer-of-training measures taken on device-trained students. In addition to this criterion-based assessment, the practical convenience of applying the models and the reliabiity of the predictive indices were also examined.

To assess the models required that a sufficient number of students be involved for whom transfer-of-training measures could be obtained as external criteria. To this end, students of four MOSs participated. The original design for generating the study data is illustrated in Figure 1. A description of analyst participants, the devices, MOSs, instrumentation, procedures, and limitations of the study design follows.

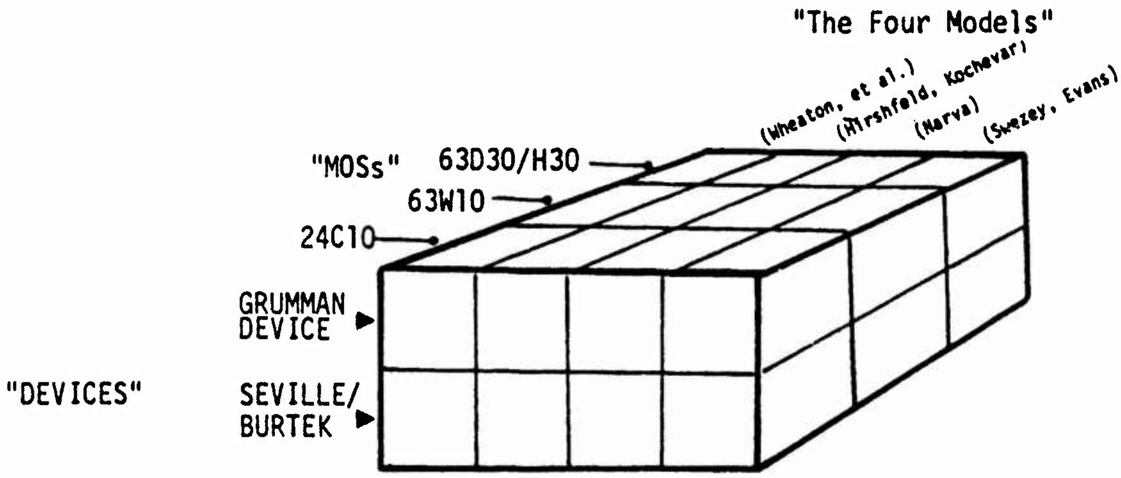


Figure 1. Study design overview

ANALYSTS (Ss)

Nine individuals served as subjects (Ss) in the study to apply the four models. Throughout this report, reference is made to this group as the "analysts" since their general assignment was to become versed in methods of the models and conduct an analysis of the training devices using each of the respective models.

Five (5) of the analysts were assigned to training devices (to be described) located at Aberdeen Proving Ground, MD. At Aberdeen, one analyst was provided by the contractor (SAI); three were Army military personnel; and one was Army civilian personnel. The SAI analyst was a researcher from the SAI Behavioral Sciences Research Center and all Army personnel were maintenance training instructors.

Four (4) other analysts were assigned to devices located at Ft. Bliss, TX. At Ft. Bliss, SAI provided a second researcher to serve as an analyst.

The Army provided one military and two civilian personnel, all of whom were qualified maintenance instructors. All of these analysts were sufficiently trained by the contractor to complete the models' protocols and serve as qualified subjects in the study.

TRAINING DEVICES

The Army Maintenance Training and Evaluation Simulation System (AMTESS) is a research and development program designed to provide the Army with cost-effective maintenance training simulators. The AMTESS concept is to provide simulators that are generic in construction and modular in design to provide flexible maintenance training at basic through advanced levels. The simulators can be modified by Army personnel for update purposes.

In the initial stage of the AMTESS effort, four different conceptual versions of the generic maintenance trainers were designed. Later, the designs of two contractors - Grumman Aerospace Corporation and a joint proposal by Seville Research Corporation and Burtek, Inc. - were selected for breadboard development. These two breadboard training devices were those to which each of the four transfer-of-training models was applied in the present study. A brief description of the devices follows.

Grumman Device

The Grumman breadboard maintenace training device is composed of six units: 1) a student CRT, 2) an instructor CRT with keyboard, 3) a desk which houses the computer and video disc system, 4) a line printer, 5) a 3-D simulation of the electrical and charging system of a diesel engine, and 6) a 3-D hawk radar transmitter simulation unit. A depiction of the device configuration is provided in Appendix A.

The student CRT (a touch screen), the instructor CRT, and the printer are located on top of a desk which houses the computer system and video disc. The 3-D units are located on a separate table a few feet away from the desk. The 3-D electrical and charging system simulation is not a life-size replica; rather, selected key components are represented. The hawk radar transmitter simulation unit is a life-size with some components absent or partly abstract in form within the interior portion of the unit.

Seville/Burtek Device

The Seville/Burtek breadboard maintenance training device consists of four components: 1) the student station, 2) an instructor station, 3) a 3-D simulation unit for a diesel engine, and 4) a 3-D simulation unit for the hawk radar transmitter. A depiction of the device configuration is provided in Appendix B.

The student station and instructor station are at separate locations in the classroom. The student station consists of a responder panel, a CRT, and a slide projector unit. The instructor station consists of a CRT with keyboard, a line printer, and a desk housing the computer system. The 3-D simulation units are full-size and are located between the instructor and the student stations. Some components of the parent equipment are selectively absent from the simulation units or are presented in partial abstract.

Both of these devices were available at each of two training sites for the conduct of the study:

- U.S. Army Ordnance Center and School, Aberdeen Proving Ground, Maryland
- U.S. Army Air Defense School, Ft. Bliss, Texas

The breadboard devices had been configured to address more than one MOS at Aberdeen, but at Ft. Bliss, they were configured to provide training for a single MOS. This did not impede the basic study design, however, since the research focused on comparing the four transfer-of-training models - the devices serving only as a testbed medium.

MILITARY OCCUPATIONAL SPECIALITIES (MOSs)

Students of four (4) MOSs participated in the study for the purpose of generating transfer-of-training measures to which predictions of the four study models could be compared. Selection of students for the study was based exclusively on "students available" at the respective training sites (Aberdeen and Ft. Bliss) during the period of field data collection - approximately May 1982 through July 1983 (study schedule to be described later). The respective number of students and MOSs participating were:

n	MOS	Description
12	63D30	Self-propelled Field Artillery Systems Mechanic
10	63H30	Direct Support Maintenance Supervisor
21	63W10	Direct Support Vehicle Repairman
20	24C10	Hawk Missile Firing Section Mechanic

Because of job similarity and task overlap, students of the 63D30/63H30 MOSs were treated as one group; criterion measures were then obtained for a set of device-trained tasks common to both MOSs. The 63W10 and 24C10 MOS students were treated as separate groups due to distinctions unique to each MOS. The result of pooling two of the MOSs produced three (3) groups of students who, for the study, served as three occupational classes on whom transfer-of-training measures were obtained (see Figure 1). The cumulative number of students fot the resulting groups was:

MOSs	n
63D30/63H30	22
63W10	21
24C10	20

Throughout the remainder of this report, transfer-of-training measures for these three occupational groups are those to which predictions of the four models will be related.

INSTRUMENTATION

Three sets of data collection instruments were used to develop study data: 1) the protocols for administering each of the four models, 2) data collection forms for obtaining the student transfer-of-training measures, and 3) the analyst opinion questionnaire (feedback on the practical utility/ convenience of applying the models). Brief descriptions of these instruments are provided below.

Data Collection Protocols

The introduction of this report discussed the four models and referenced the literature describing each. Each model possesses its own set of procedures for generating data required to produce a transfer-of-training forecast. These procedures are lengthy, relatively complex, and generally differ with each model. Their description is beyond the scope of this report and thus the reader is referred to the source documents for complete reviews of the protocols. One parital example of the analyst's worksheet and computations, taken from the Swezey and Evans (1980) model is provided in Appendix C. This worksheet and comparable materials for the remaining models were assembled in a "package" and provided to each analyst.

Transfer-of-Training Measures

Percent of steps passed on equipment operations served as the transfer-of-training measure. This was taken on students following device training; however, the performance test forms utilized by the schools were found to be too global in nature for effectively assessing transfer-of-training. Therefore, detailed data collection forms were developed for each training task (to be described later) on which transfer-of-training measures were taken. These forms allowed the data collector to obtain information such as:

- Student identification
- Dichotomous GO/NO GO data for each task step (source of the criterion measure for this study)
- Comments or relevant details on the student or testing situation

These instruments were developed for the respective MOSs by consulting technical manuals and subject matter experts to determine appropriate content. Preliminary versions of the forms were then pilot-tested and refined accordingly. The test forms served the needs of both the present study and the larger AMTESS research effort. A copy of one revised test for the 63W10 MOS students is provided in Appendix D. A complete review of these instruments is provided in Unger, Swezey, Hays, & Mirabella (1984).

Analyst Opinion Questionnaire

The analyst opinion questionnaire was administered after analysts had completed all applications of the models to the devices. The questionnaire was designed to obtain analyst feedback on the practical value and convenience of applying each model. Specifically, the self-administered survey asked the nine analysts how they felt about the models in terms of effectiveness and user difficulty. A copy of the analyst opinion questionnaire is provided in Appendix E.

PROCEDURE

Study Schedule and Location

The study was executed as a subordinate component of a larger research effort evaluating two AMTESS training devices (the Grumman and Seville/Burtek devices described earlier). The schedule for the present study thus proceeded concurrently with that of the larger AMTESS effort. The schedule is depicted in Figure 2, indicating periods when the two devices were installed/accepted at their respective locations through the time of data collection/evaluation. As pointed out earlier, the reader will note from Figure 2 that the two devices were available at both the Aberdeen and Ft. Bliss training sites where this study was conducted.

Preparation of Subjects

Of the study subjects (the nine analysts), five were assigned to the Aberdeen site and four to Ft. Bliss. The SAI field researchers, assigned to Aberdeen and Ft. Bliss, were responsible for familiarizing all other subjects at their sites with the four respective models, their methods and materials. This was accomplished and all subjects were provided with protocol materials for completing the four analyses on each of the Grumman and Seville/Burtek devices.

Application of the Models

The Grumman and Seville/Burtek devices were capable of training a variety of MOS tasks; however, to maximize the number of student participants on whom transfer-of-training measures could be obtained, only a limited number of training tasks were studied. These were procedural tasks common to certain MOSs, thus permitting more students to be involved. The analysts (Ss) were made aware of the maintenance procedures to be assessed and were instructed to apply each of the four models (to the device) for only those procedures [3]. For the respective devices, MOSs and training sites, the procedures to which the models were applied are summarized in Table 1.

---

[3] Each procedure was a maintenance "procedure" and consisted of several tasks/subtasks which were the actual focus of the application of the four models. A complete description of these tasks/subtasks is provided in Appendix F.

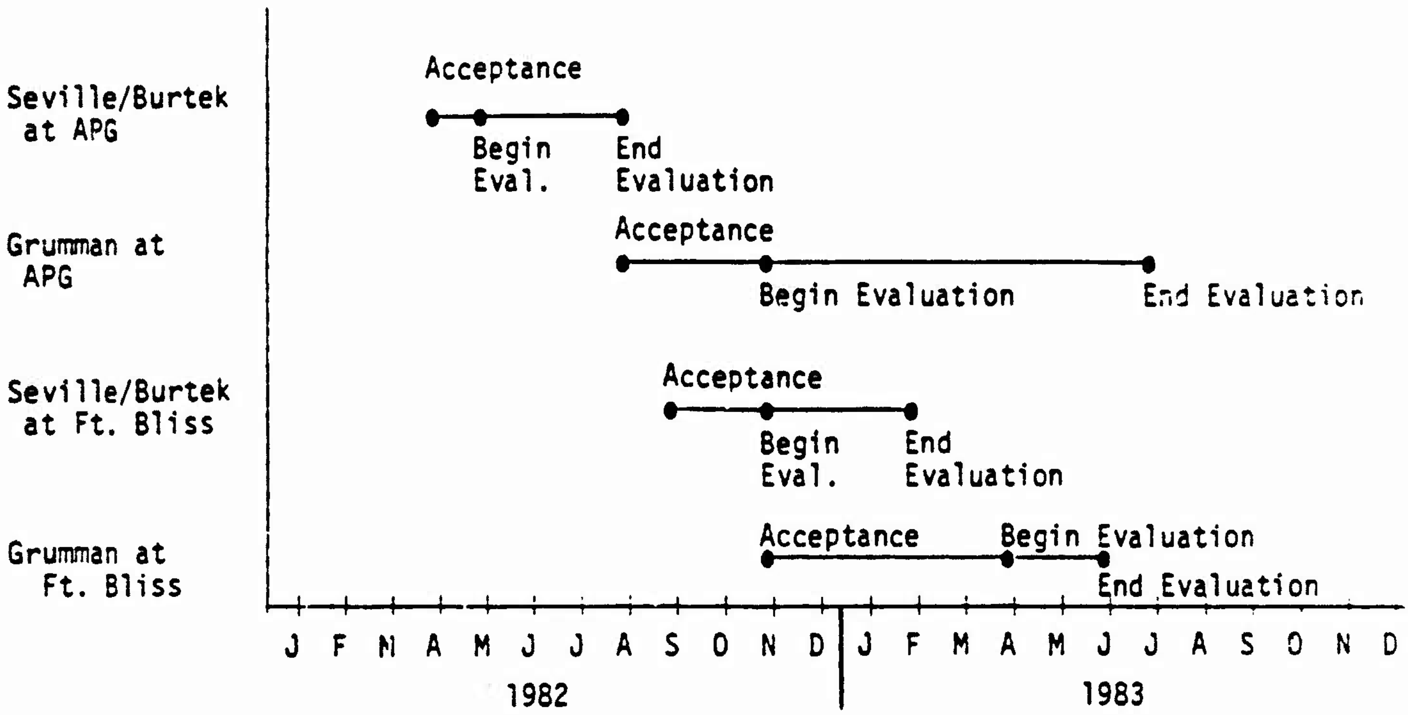


Figure 2. Schedule for the concurrent studies

TABLE 1

PROCEDURES ON WHICH PREDICTIONS AND TRANSFER-OF-TRAINING MEASURES WERE OBTAINED (by site, MOS, and device)

SITE	MOS	DEVICE: Grumman	DEVICE: Seville/Burtek
ABERDEEN	63D30/63H30	1) Start engine, confirm generator warning light on 2) Perform VTM hook-up and check-out	--
	63W10	--	1) Troubleshoot engine malfunction 2) Replace oil pump filter and pump release
FT. BLISS	24C10 (same tasks for each device)	1) Weekly check procedure for hawk radar transmitter	

Because applying the four models was a labor-intensive and time-consuming effort, it was not possible for all analysts to evaluate both devices. Rather, each analyst was assigned to one device at their site - although some analysts assessed both devices. Every analyst, however, did apply all four models to the device(s) to which they were assigned. The number of complete assessments obtained per device is shown in Table 2. Once the assessments were completed, four (4) predictions (one for each of the four models) were available from every analyst indicated in Table 2.

TABLE 2

NUMBER OF ANALYSTS (Ss) ASSESSMENTS
(by device and site)

SITE	GRUMMAN DEVICE	SEVILLE/BURTEK DEVICE
ABERDEEN	Ss = 3	Ss = 4
FT. BLISS	Ss = 2	Ss = 3

Analyst Feedback

After completing application of the models, each analyst was provided with the analyst opinion questionnaire to obtain his/her view on the practical utility/conventience of implementing the models.

Transfer-of-training Measures

For each MOS student participating in the study, task transfer-of-training measures were taken within 24 hours of completion of device training. Instrumentation used to obtain the measures was described previously, one example of which is provided in Appendix D. These tests were administered at Aberdeen and Ft. Bliss by the respective on-site SAI researchers.

Data Analysis

Figure 1 of this chapter illustrated the general design for the study in terms of MOSs and devices to which the four mdoels were applied. Later, Tables I and 2 (respectively) pointed out that not all of the MOSs were trained by each device, and that the number of analysts-as-subjects differed per device. These deviations from the Figure 1 design were generally known in advance of the study. For example, the researchers were aware that the devices had been set up at the schools with the intent of training different MOSs. On the other hand, variation in the number of analysts per device was not predicted. Rather, that variation was situationally induced due to unexpected inavailability of Army personnel who intended to serve as analysts. The net effect of these occurrences is shown in Figure 3, which illustrates (by darkened areas) those cells for which study data could be generated and the number of analysts (Ss) who applied the four models. The data produced for each darkened cell, therefore, were those subject to data analysis in the study. A number of limitations burdened the study data, the most significant of these being the small number of analysts (Ss) available to apply the four models. (These limitations will be discussed later in this section). Suffice to say that in light of these limitations, the study should be regarded as a "preliminary estimate" of the criterion validity and reliability of the respective models although analyst opinion regarding practical utility of the models may hold somewhat more immediate value.

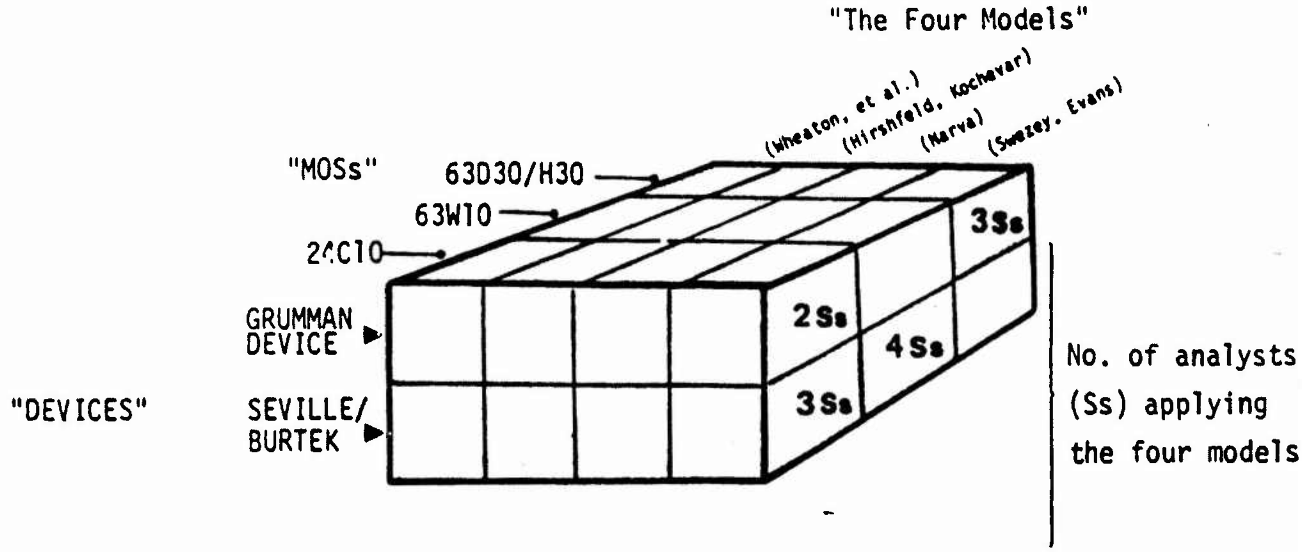


Figure 3. Final study design (study data available from dark cells only)

Data analysis assumed an hypothesis of no difference between models and compared them on the following dimensions:

- Predictiveness - the relationship between actual model prediction and transfer-of-training measure
- Reliability - inter-analyst agreement (based on different scores) for all predictions
- Practical utility - analyst's perceptions of the ease-of-use and effectiveness of the respective models

Due to the small number of analysts-as-subjects, data analysis compared the models through correlation methods supplemented by descriptive statistics. Inter-analyst agreement was computed as an index of reliability, and analyst opinion regarding each model was summarized in terms of opinion ratings and open-ended narrative comment. The findings are documented in Section III of this report. Before introducing the findings, however, it is appropriate to describe certain study limitations which affect the generalizability of the results.

STUDY LIMITATIONS

In reviewing study results, the reader should keep in mind a number of constraints imposed on the study. The extent to which these potential confounds may have had effect was not determinable since they were either

unanticipated or could not be controlled through available time and study resources:

- Number of subjects. As noted previously, the number of analysts-as-subjects was small and does limit the generalizability of study results. A repeated-measures-within-subjects approach was used to analyze the data and help compensate for this limitation.

- Ceiling effects on the criterion test. Inspection of student test results for the transfer-of-training measure suggested the possible presence of ceiling effects although this could not be satisfactorily confirmed.

- Narrow range of training tasks. Training tasks selected for inclusion in the study were few. These were determined by overlap between training offered by the devices and actual MOS training needs. Possibly performance of the models could vary across a broader range of tasks.

- Participant cooperation. Cooperation of participants in the study varied. Generally, most were cooperative. However, some students expressed disinterest in participating and some lack of Army administrative support was evident at the training sites. Also, some TRAINVICE analysts were disgruntled about the time and application difficulty of the models. What impact, if any, this had on study results is not known.

- Setting interference. A number of setting interferences occurred. Although any one was slight, their cumulative effect is possibly of some concern. These interferences were:

  - Due to student schedules and limited accessibility, it was not possible to take pre-measures on the MOS students.

  - Training devices often broke down; down-time ranged from one hour to a week. This interrupted both training and application of the models.

  - Student transfer-of-training tests were sometimes interruped by failures or unavailability of operational field equipment.

  - Student training and testing was sometimes interrupted by superceding activities (e.g., unannounced fire drills, students leaving to participate in some other duty).

  - At Aberdeen some instructors rotated, thus providing some students with multiple trainers.

III. RESULTS

ORGANIZATION OF DATA

Data generated by the study were of three categories:

1. Each model's transfer-of-training predictions for the devices

2. Student transfer-of-training measures (% steps passed) for each MOS

3. Analyst feedback (questionnaire results) on practical utility of the models

All computations generating these data were reviewed for completeness and any missing entries replaced with best estimates based on data produced by other subjects. Since only two to four analysts were available per device, averages were used to estimate missing values if one analyst produced missing data but at least two other analysts' protocols were complete. Otherwise, a single analysts' data was used to estimate missing values when only two analysts had been assigned to a device and one analysts' protocol showed missing entries. There was no alternative to this approach. Fortunately, estimating missing values proved necessary only in the case of two subjects' administrations of a training techniques analysis.

Setting aside the data collection category concerning analyst feedback for the moment, data for the first two categories above were organized in the following manner. First, a data matrix was constructed to record all analyst-generated transfer-of-training predictions. These data were arranged by training tasks within analysts within models. The matrix contained prediction indexes at the "task" level and also for the "summary index" which each particular model generated as its terminal figure of merit. These data were available on all analysts for every model application that they conducted. In addition, inter-analyst agreement was computed for each application. Also, student transfer-of-training (ToT) measures were averaged to produce criterion means and paired with predictions at both the task level and summary index levels. The data organization scheme is illustrated in Table 3.

Study results were determined for task level and summary predictions and analyst opinion was reviewed in light of the findings. The following subsections report results of the data analysis for each model's summary predictions, task-level predictions, reliabilities, and analyst feedback respectively.

SUMMARY PREDICTION

Each model produces a summary "figure of merit" which serves as the terminal prediction of device ToT potential. This summary index, for all four models, ranges from 0 to +1 such that as the index approaches +1, the greater the ToT of the device is presumed to be. In the case of all models, the summary index is essentially the average of the individual effectiveness predictions made on each training task trained by the device.

TABLE 3

ORGANIZATION OF DATA

For application of model "X" to MOS "Y" (device "Z"), the available data were:

	MODEL "X"			OTHER DATA PAIRED TO PREDICTIONS	
Analysts:	S_1	S_2	S_n	Inter-analyst Agreement	Criterion (ToT) Measure
Task-level predictions on training tasks:	Task a	T_a	T_a	r_a	$\bar{x}_a$
	T_b	T_b	T_b	r_b	$\bar{x}_b$
	T_c	T_c	T_c	r_c	$\bar{x}_c$
	T_n	T_n	T_n	r_n	$\bar{x}_n$
Summary Prediction:	P_{S_1}	P_{S_2}	P_{S_3}	r_P	Grand mean (ToT)

Use of the summary prediction index alone was not sufficient for purposes of validating the models. This was due to the fact that no more than two to four analysts produced each summary index and thus the number of cases available was too limited to draw definite conclusions on the basis of the summary index alone. More importantly, the summary index is a "product" of many subordinate analyses conducted within each model at the task level. Thus, the summary index is valid (i.e., not an artifact) only to the extent that each task-level prediction which generates it is valid. These task-level predictions thus became the main focus of the study. Still, it is appropriate to present the summary prediction outcomes first since later findings will be related to them. The summary predictions are reported in Table 4.

From the data in Table 4, it appears that summary predictions of the Hirshfeld and Kochevar model tends to covary more with the criterion than do the other models. This is by no means a conclusive finding, however, since the summary index could well be a misleading artifact of task level predictions (i.e., perhaps task-level predictions which form the summary figure are not valid), or the summary index could be subject to malfunctions of model mathematics. To seek the root of the validity question, a more critical

TABLE 4

ANALYSTS' (Ss) SUMMARY PREDICTIONS AND STUDENT ToT MEANS

	(Ss[1])	Summary Predictions: Wheaton	Hirshfeld	Narva	Swezey/Evans	Mean ToT (% steps passed)
ABERDEEN:						
• Grumman device (MOS 63D/H)	S_1	.45	.94	.28	.81	91.5%
	S_2	.31	.91	.28	.82	
	S_3	.50	.94	.21	.85	
• Seville/Burtek (MOS 63W)	S_1	.21	.81	.19	.81	68.0%
	S_2	.31	.75	.10	.57	
	S_4	.32	.85	.44	.79	
	S_5	.43	.77	.19	.74	
FORT BLISS:						
• Grumman device (MOS 24C)	S_6	.46	.98	.34	.77	92.8%
	S_7	.47	1.00	.05	.79	
• Seville/Burtek (MOS 24C)	S_6	.25	.98	.38	.84	93.8%
	S_8	.27	.95	.28	.86	
	S_9	.26	.99	.30	.91	

[1]Note: Some subjects applied the models to more than one device and thus appear twice in the table.

analysis of the data was conducted on the task-level level predictions.

TASK-LEVEL PREDICTIONS

As noted earlier, the summary prediction of each model is actually a function of the particular model applied to each and every task trained by the device. Some of the models conduct their analysis of a device at the task or subtask level while others assess the skills/knowledges involved in the trained tasks. It is only at the "task" level, however, where all of the models produce an index that can be compared across models. It should be noted that this index is derived from calculations of each model as a "sub-step" in the mathematical process to achieve the summary index. The task-level index was not put forth by originators of the models as a terminal metric. The task level index was, nonetheless, the most discrete level of analysis which could be undertaken to study predictive validity of the models, and was also of interest for the following additional reasons.

First, it is conceivable that two analysts applying the same model to the same device could produce opposing views of the device's effectiveness. One analyst might rate the first half of the device-trained tasks favorably and the remaining half as ineffective training. The second analyst could take just the opposite view. Yet, because of the summation process used to produce the terminal index, both analyses might result in the same or at least a very comparable summary prediction.

A second reason for assessing model predictions at the task level concerns the untested mathematics of the models. Chapter I of this report noted that at least parts of the models' mathematics were questionable and that some prediction error was probably inherent due to weaknesses in those formulations. This error may be slight for individual tasks assessed by a model, but may accumulate across the summation process to distort the summary index. Comparing model predictions to student ToT measures at the "task" level provides a more discriminating analysis which effectively partitions meaningless variance that might be present as a function of summation methods of the models.

To assess the models using the task-level predictions as the comparison basis, the following steps were taken. First, for each model, all analysts' task-level predictions were listed for the MOS tasks trained. This provided a total of 34 predictions developed from each model. Second, the criterion measure mean for each trained task was determined from student ToT scores and was listed alongside its corresponding task-level prediction. This produced 34 paired cases - each case including a task-level prediction and a respective criterion (ToT) measure. The organization of data for one model is illustrated in Table 5. (The actual data on which analysis was conducted, arrayed for each of the four models, is given in Appendix G). For each model, the column of task-level indexes in Table 5 was correlated with that of the paired criterion measures to determine the strength of relationship between task-level predictions and ToT. The Pearson product-moment correlation was employed for this purpose. As Table 5 and Appendix G indicate, the paired criterion measure "repeats" for analysts' task-level predictions on each particular device-trained task. With the limited number of analysts and training tasks available, there was no alternative to this approach. This did not prove to hinder analysis, however.

TABLE 5

DATA ORGANIZATION FOR COMPARING TASK-LEVEL PREDICTIONS TO ToT CRITERION

	Analysts[1]	Task-level predictions for model "n"	Criterion (mean student ToT scores for each task)
DEVICE X:			
Task 1	S_1	$t_1(S_1)$	$\bar{X}_{t_1}$
	S_2	$t_1(S_2)$	$\bar{X}_{t_1}$
	S_3	$t_1(S_3)$	$\bar{X}_{t_1}$
Task 2	S_1	$t_2(S_1)$	$\bar{X}_{t_2}$
	S_2	$t_2(S_2)$	$\bar{X}_{t_2}$
	S_3	$t_2(S_3)$	$\bar{X}_{t_2}$
DEVICE Y:			
Task n	S_1	$t_n(S_1)$	$\bar{X}_{t_n}$
	S_2	$t_n(S_2)$	$\bar{X}_{t_n}$
	S_4	$t_n(S_4)$	$\bar{X}_{t_n}$
	S_5	$t_n(S_5)$	$\bar{X}_{t_n}$
etc.			

[1]NOTE: Recall that some analysts served to evaluate more than one device. This is illustrated above for device "Y", which Ss #1 and #2 evaluated in addition to device "X".

Alternative data sets of non-repeating criterion measures were configured and tested against the results derived from correlating the repeating criteria with task-level predictions. Use of the repeating criteria proved inconsequential. The correlation results for each model in the study are provided in Table 6.

Table 6 presents a strikingly different picture than presented earlier in Table 4 (summary index results). In Table 6, the Hirshfeld and Kochevar model and the Narva model appear non-predictive of the criterion: their correlations with the criterion failing to reach significance. On the other hand, the original Wheaton et al. model, and the Swezey and Evans model do correlate positively, though modestly, with the criterion; both correlations being comparable and significant at $p<.05$.

The data suggest that these latter two models possess some predictive potential. Ostensibly, their correlations are modest although these correlations may be depressed due to quirks of model mathematics or possible ceiling effects present in student ToT measures. This is presently not verifiable, however, and must be ascertained through future research. It does appear from this analysis that the models summation processes do produce a misleading summary index (see Table 4 in comparison to Table 6) and that the summary index should not be relied upon to make device selection decisions. Rather, such decisions should be made using direct comparison of task-level indexes for competing devices.

INDEX RELIABILITY

Since two or more analysts applied each model to the various MOSs, it was a simple matter to determine inter-analyst agreement as an estimate of reliability. Reliability of agreement is the appropriate measure when absolute concurrence among Ss is the desired circumstance, as is the case with the four models of this study. The limited number of analysts did preclude the use of more preferred regression estimates of reliability such as the intra-class correlation coefficient (Shrout and Fleiss, 1979). Instead, the reliability estimate was based on difference scores between all possible pairs of Ss and reflects "percent of agreement". The agreement estimate (p_a) was calculated by first determining the proportional "value" of a single interval on the task-level index scale of .00 to 1.00. Difference scores for each pair of analysts were then determined and multiplied by that value. The resulting scores for all pairs who assessed the particular task were then averaged and that mean subtracted from 1.00 to produce the mean percent of agreement. Thus, the estimate should not be construed as a correlation coefficient, but stands, nonetheless, as an appropriate estimate of reliability for study purposes.

Reliability of agreement was determined for each model for both the task-level and summary indexes. Because only one summary index is generated for a particular model, inter-analyst agreement at the summary level was calculated directly from the summary index by averaging it for all analysts' administrations of the model. At the task-level, however, inter-analyst agreement was calculated for the task index of each task, then averaged across tasks for all administrations of the model. The standard deviation of the agreement mean was determined for both reliability estimates. Results of this assessment are presented in Table 7.

TABLE 6

CORRELATIONS (r) BETWEEN TASK-LEVEL INDEX AND CRITERION (TASK-LEVEL ToT)

	Wheaton, et al.	Hirshfeld, Kochevar	Narva	Swezey, Evans
r (Validity Coefficient)	.33*	-.03 n.s.	.07 n.s.	.34*
r^2 (Proportion of Variance Accounted for)	11%	.1%	.5%	12%

NOTE: * Significant at $p<.05$
Correlations are based on 34 paired cases for each model.

TABLE 7

ANALYST AGREEMENT FOR THE

TASK-LEVEL AND SUMMARY

PREDICTION INDEXES[1]

	Wheaton, et al.	Hirshfeld, Kochevar	Narva	Swezey, Evans
Task-Level Index	.91 sd=.11	.83 sd=.14	.83 sd=.14	.95 sd=.05
Summary Index	.94 sd=.06	.97 sd=.02	.86 sd=.11	.94 sd=.05

1 In all cases, reliabilities given are the "mean percent of agreement" among analysts as determined from the sum administrations of each model.

Table 7 shows that analyst agreement on both task-level and summary predictions tends to be high. There is a tendency for summary prediction indexes to reflect slightly higher agreement. Since task-level index agreement derives from judgment activities close to model raw data, however, it is likely the safer estimate of reliability. The elevated reliability of the summary predictor, on the other hand, could be an accumulation of reliable, but contaminating, variance due to peculiarities of the respective model's summation procedures (mathematics). This possibility will be discussed later.

PRACTICAL UTILITY

The amount of time required to administer the four models to a single device was reported to be approximately three intensive eight-hour days (in some cases, spread over a seven-day period). Following administration, the nine analysts were asked to provide feedback on their experience in applying the models. Feedback was in the form of unrestricted narrative comment and ratings of the models regarding their effectiveness and administration difficulty. (Note: Because the Narva model and the Swezey and Evans model are similar and analyst tasks essentially the same, a single set of questions was provided to the analysts to obtain feedback for those two models). Narrative comments were as follows (ratings are reviewed separately):

- <u>Wheaton et al. Model</u>

  - <u>Training techniques analysis</u> was extremely difficult to understand and administer; terminology is beyond most analysts' level; rating scales are difficult to use and are repetitive (punishing) to rate; takes too long to administer; terminology is too "jargonish" and inappropriate for those expected to apply it; Learning Guidelines don't make sense for some aspects of the device. (Note: Comments on the training techniques analysis were by far the most prevalent for this model).

  - <u>Structure of the model</u> is too detailed/complicated; sections are tedious to complete due to many judgements; poor method for assessing simulator effectiveness; model should be restructured (four respondents).

  - <u>Validity of judgments</u> is questionable; requires a highly qualified person to do; analyst must take many breaks to produce high-quality ratings (two respondents).

  - <u>One analyst</u> reported that he liked the model and had no criticism of it.

- <u>Hirshfeld and Kochevar Model</u>

  <u>Very few comments</u> were made by the analysts on this model although the following comments were provided by five respondents:

  - <u>Easiest model</u> to apply; liked the model (two respondents).

  - <u>Poor method</u> for assessing device effectiveness; takes too long to complete; punishing for analyst to apply; involves too much

technical terminology/jargon; requires a highly qualified person to do; task training difficulty analysis is probably least accurate component due to heavy reliance on analyst judgment; model should be restructured (three respondents).

- Narva; Swezey and Evans Models

  - Physical/functional characteristics analysis was too difficult, especially use of the Learning Guidelines and behavioral categories; very hard to make judgments; too much material and reading required in this analysis; physical characteristics/functional characteristics requires extensive time to complete; too much technical jargon involved; too detailed; examples in the Learning Guidelines don't apply to the device; too cumbersome to apply. (Note: Comments on the physical/functional characteristics analysis were the most prevalent for these two models).

  - Generic characteristics list was useful/sensible; generic characteristics list was easy to use and a more objective format than remainder of physical/functional characteristics analyses (two respondents made these comments).

  - Models are too difficult to have any true application; training proficiency analysis was difficult; models require a very qualified person to apply; too detailed/jargonish; vocabulary is inappropriate for those expected to apply the model (six respondents).

Analysts Ratings of the Models

Analysts were asked to rate each component of the models regarding: 1) difficulty in applying each component of the model, and 2) effectiveness of the judgmental data (device measures) which the analysts produced. The results of this feedback are presented in Tables 8 and 9, respectively, showing means and standard deviations for the analysts' responses.

Of the models which earlier proved to be the more predictive (Wheaton et al.; Swezey and Evans), Tables 8 and 9 show those models to have fairly similar profiles. Notably, the Training Techniques Analysis of the Wheaton et al. model and the Physical/Functional Characteristics Analyses (the latter actually a training techniques analyses) of the Swezey and Evans version were viewed as the most difficult to administer (Table 8) and most ineffective with respect to analyst judgment (Table 9). For all four of the models, analyses associated with "learning deficit" (i.e., skill/knowledge requirements, task training difficulty, training proficiency, learning difficulty) were viewed as somewhat less difficult and less ineffectual. Remaining model componenets concerning task communality, similarity, and coverage requirements were judged to be relatively easy to apply and more effective with respect to device assessment. The Hirshfeld and Kochevar model was viewed as easier to apply than the others, but essentially no more effective.

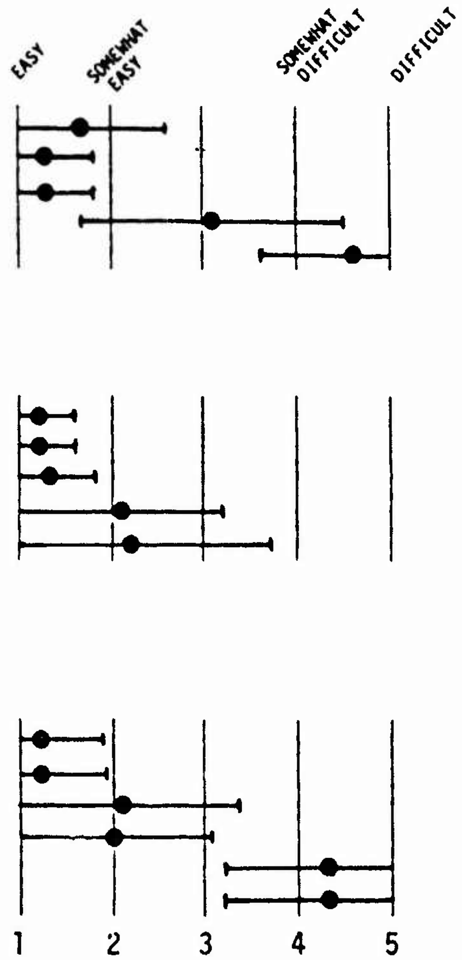
One remaining finding becomes strikingly obvious if the reader compares the visual appearance of Table 8 data to that of Table 9. Specifically, the two tables are virtually "mirror images" of one another. Regarding this,

TABLE 8

ANALYSTS' "DIFFICULTY" RATINGS

FOR THE FOUR MODELS

MAKING DEVICE JUDGMENTS
(RATINGS) WAS:



Wheaton et al. Model

- Task communality
- Physical similarity
- Functional similarity
- Learning deficit
- Training techniques

Hirshfeld & Kochevar Model

- Task communality
- Physical similarity
- Functional similarity
- Skill/knowledge requirements
- Task training difficulty

Narva; Swezey & Evans Models, respectively

- Coverage requirements
- Coverage
- Training proficiency
- Learning difficulty
- Physical characteristics
- Functional characteristics

Legend:

● = mean ($\bar{x}$) rating
├●┤ = ± 1 standard deviation
No. cases = 9 analysts

TABLE 9

ANALYSTS' "EFFECTIVENESS" RATINGS

FOR THE FOUR MODELS

IN ASSESSING THE DEVICE, JUDGMENTS WERE:

Wheaton et al. Model

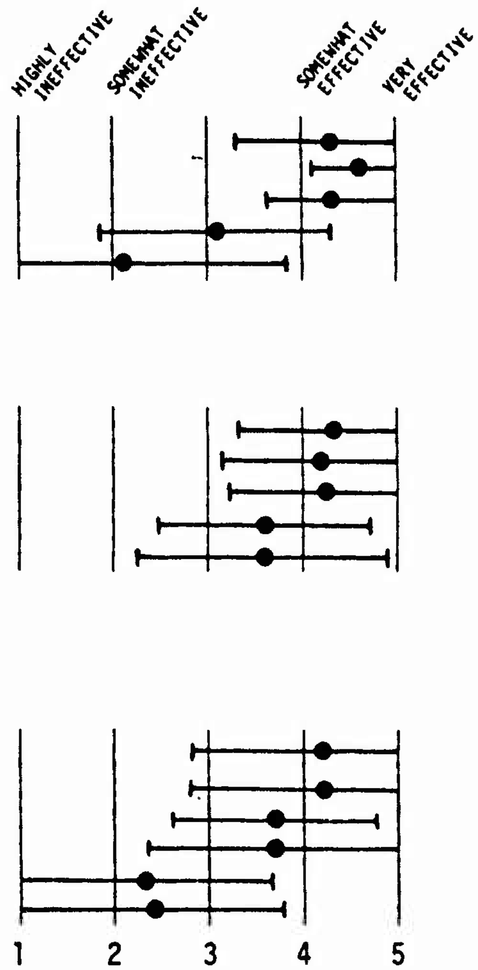
- Task communality
- Physical similarity
- Functional similarity
- Learning deficit
- Training techniques

Hirshfeld & Kochevar Model

- Task communality
- Physical similarity
- Functional similarity
- Skill/knowledge requirements
- Task training difficulty

Narva; Swezey & Evans Models, respectively

- Coverage requirements
- Coverage
- Training proficiency
- Learning difficulty
- Physical characteristics
- Functional characteristics



Legend:

● = mean ($\bar{x}$) rating

├─●─┤ = ± 1 standard deviation

No. cases = 9 analysts

the analysts' feedback is clear: simply, the easier the model was to administer (in the analysts' opinion), the more effective it was perceived to be. This position is not well-supported by task-level index findings since the easiest model to apply (Hirshfeld and Kochevar, according to the analysts) proved to be the least predictive model in the study. Nonetheless, any prediction method should seek ease and effectiveness in application and this preference is born out by the analysts' feedback. If this is not accomplished eventually, the models will be useful only to highly trained/motivated specialists and denied to a broader range of personnel.

ACCOUNT OF VARIANCE AND SUMMARY

Based on results of the data analysis, an accounting of each model's variance was determined. This accounting, along with a summary of all other findings, is presented in Table 10. The reader should recall that the predictive validity of each model was ascertained from each model's task-level index and all figures in Table 10 derive from that level of data analysis. The accounting of variance is based on the following convention (see: Kerlinger, 1973; Harmon, 1967; or Nunally, 1978):

$$V_T = V_c + V_u + V_e$$

where:

V_T = total variance

V_c = common factor variance (validity)

V_u = unique variance (reliable measurement contamination)

V_e = measurement error variance

and:

$$V_c + V_u = \text{reliability}$$

The reader should be aware that V_c, V_u and V_e each represent squared quantities. For example, validity (V_c) is derived in this study as the square of the validity coefficient (r)which results in r^2 - the proportion of variance accounted for in common by the predictor variable and corresponding criterion variable. Further, it is important to understand that V_c and V_u are "reliable" variances and when combined must equal the reliability. Reliability is thus assumed to be the sum of these two squared entities and is, itself, never squared in accounting for variance. Rather, reliability serves as an important reference figure which permits the calculations to be made which provide a full accounting of variance. These principles are reflected in Table 10 data organization.

TABLE 10

SUMMARY OF FINDINGS

- SUMMARY INDEX RESULTS

This index proved to be a misleading predictor, apparently accumulating irrelevant, yet reliable, variance through the summation process. Possibly, quirks of each model's mathematical formulation may encourage this distortion of the terminal prediction at least in part.

- TASK INDEX RESULTS

	Wheaton, et al.	Hirshfeld, Kochevar	Narva	Swezey, Evans
<u>Validity Coefficient</u> (r) (correlation between prediction and ToT criterion)	.33 $p<.05$	-.03 n.s.	.07 n.s.	.34 $p<.05$
<u>Reliability</u> (p_a) (analyst percent-of-agreement on the task-level index averaged across training tasks; used here as the reliability estimate)	.91 sd=.11	.83 sd=.14	.83 sd=.14	.95 sd=.05

- ACCOUNT OF VARIANCE

	Wheaton, et al.	Hirshfeld, Kochevar	Narva	Swezey, Evans
<u>Validity</u> (r^2)	11%	.1%	.5%	12%
<u>Measurement Contamination</u> ($p_a - r^2$) (undefined but reliable variance)	80%	82.9%	82.5%	83%
<u>Error</u> ($1-p_a$)	7%	7.0%	7.0%	5%
TOTALS	100%	100%	100%	100%

- ANALYST FEEDBACK

Analysts-as-subjects viewed the models as difficult and time-consuming to administer, feeling that the models should be restructured. Training techniques analyses were seen as the most cumbersome. In their opinion, the easier the model was to administer, the more likely it was effective.

IV. CONCLUSIONS AND RECOMMENDATIONS

CONCLUSIONS

This study compared the efficacy of four transfer-of-training prediction models; employing two forms of training devices and three categories of maintenance MOSs as the test-bed. The scope of the study was limited and constrained by a number of possible confounders described in Chapter II. Findings, therefore, should be regarded as preliminary estimates of the validity and reliability of the models. At the very least, this study provides future researchers with hypotheses regarding metric properties and practical utility of the models and describes potential threats to research control which might be encountered if replicating in a comparable setting.

Based on the data generated in the study, the results of analysis permit the following conclusions to be drawn:

1. The summary index of the models is a misleading predictor. This index appeared to be a distorted metric. When the predictive power of the models was assessed at the task level, their efficacy proved considerably different from that suggested by the summary index. Because the mathematics of each model are questionable in many respects, and because the task-level and summary indexes remain reliable throughout computation, the cause of this distortion is possibly a malfunction of mathematics in large part.

   Since the same math model (for any one model) produces both the summary and task-level index, presumably distortion is present in the task-level index also, but to a lessor extent than at the summary level. As the task-level indexes are summated to produce the summary index, distortion likely accumulates over the many subtasks (or skills/knowledges) involved to produce a misleading terminal prediction. For the present models, the terminal summary index should not be used to predict device effectiveness.

2. Comparing the transfer-of-training potential of competing devices should rely upon task-level predictions. As a corollary to the above finding, any comparison of devices through application of the present models should rely upon the task-level index. This would require comparing the devices on a "training task-by-training task" basis and not relying upon a single, summary figure of merit to represent each device's overall transfer potential. In so doing, the more valid models should be used to generate the task-level indexes.

3. The Wheaton et al. and Swezey and Evans models are the more predictive models. These proved to be the more valid models in this study although their correlation with the criterion was modest at .33 and .34, respectively. All things con-

sidered, the Swezey and Evans model appeared to be the slightly more efficient of the two. The Hirshfeld and Kochevar and the Narva models were not predictive of the criterion, producing essentially "zero" correlations with the ToT measure.

4. Both the task-level and summary indexes proved to be highly reliable. Agreement between analysts was high for both task-level and summary predictions -- ranging from the .80s through .90s. This should not be taken to suggest that analyst judgmental ratings are consistently as reliable. (Conceivably, the mathematical models which produce the indexes could be insensitive to minor variation in analyst agreement at the raw data level). The high index reliabilities did make clear, however, that the models possess much unexplained yet reliable variance (from 80% to 83%). Possibly, this contamination is a function of mathematical quirks of the models or the measurement of something other than ToT potential - or possibly some of both.

Discussion: Utility of the Models

The practical utility of the four models, in light of findings here, must be considered from two perspectives: 1) the measurement viewpoint, and 2) that of the field user. From the measurement view, it is always disappointing to find only modest validity coefficients produced by the superior tests. This was the case in the present study; the Wheaton et al. and Swezey and Evans models (the best predictors at the task-level) correlated only .33 and .34, respectively, with the criterion. Giving benefit of the doubt for the moment, it is possible that these correlations could actually be higher if ceiling effects are present in the criterion test. Ceiling effects, in departing from linearity, would make each model's correlation with the criterion smaller than its correlation with true performance and thus underestimate the value of "r". Attenuation of this type due to errors of measurement can be corrected for a truer estimate of r (Carmines and Zeller, 1979). In the present study, however, this was precluded due to limitations on study data needed to calculate such adjustments. Future research should consider this possibility although it is doubtful that adjusted correlations will show more than modest gains -- the better control would be the use of highly discriminant criterion tests.

The prediction indexes of the models were highly reliable. However, this asset is presently diminished in light of the large proportion of contaminating variance each model possesses. Future improvements in the validity of the models may, in fact, reduce their reliability. However, it would be much better to see 70% of the variance constituting validity, at the expense of some reliability, than the present case. With respect to these psychometric considerations, one fact is abundantly clear. In order to make optimal utility of the two superior models, it is essential that devices be compared on a "task-by-task" basis using the task-level index. The summary index is simply too misleading.

On the basis of the findings, one is inclined to conclude that even the two best models (Wheaton et al.; Swezey and Evans) prove to be weak predictors of ToT potential. What is not known, however, is: How accurate

are device ToT predictions in the absence of the use of these better models? Perhaps unaided approaches produce even less efficient decisions as to training device effectiveness. Such judgments, aided by one of the more predictive models, might improve otherwise unaided decision accuracy. Until statistics on the accuracy of unaided ToT predictions become available, it is not possible to determine the practical utility of the existing models in cost-savings terms.

Last, we are reluctant to conclude that the Hirshfeld and Kochevar and the Narva models be discounted from future research because of their performance in this study. All four of the models presently remain embryonic in development and may be victims of mathematical modeling errors, faulty inclusion or exclusion of construct variables which confound prediction, or minor problems that introduce irrelevant variance. Future research should make the attempt, therefore, to go beyond studying prediction "indexes" and determine the contribution which each "variable" (i.e., learning deficit analysis, PC/FC analyses, etc.) makes to prediction. Only that level of investigation will bring about the insight necessary to understand why each model works effectively or why it does not.

The other perspective to be considered in judging practical utility is the view of the field user -- the individual who must "apply" the model and rely upon its results. In the present study, nine TRAINVICE analysts assumed this role. Although the analysts were not in a position to comment on psychometric properties of the models, they did provide feedback on the practical convenience of applying the models. Generally, this feedback was extremely negative; the models being seen as too time-consuming, cumbersome, technically complex and discouraging to be of practical value.

From the early days of TRAINVICE R&D, this problem was anticipated. Certainly feedback from analysts in this study merely confirmed this suspicion. Psychometrists have long recognized that no matter how valid the test, if it is impractical to implement, then its utility is comparably diminished. In large measure, the various models suffer from the problem of unwieldliness in application and scoring; the seriousness of this problem appears to be of major proportion. Current and future research efforts should strive to develop simpler models to comprehend, administer, score, and interpret.

RECOMMENDATIONS

1. The present study should be replicated to increase the level-of-detail of investigation and control; thereby producing definitive findings on the efficacy of the models. The following should apply to the research design:

    a. All four models should be retested including any new advancement in model design that might be developed.

    b. The research setting should be highly controlled and conducive to obtaining accurate measures of each model's metric properties. The university laboratory is preferred over constraints of the military field setting.

c. The subject pool should be increased to 15-25 analysts. Paid graduate students in psychology or educational technology would be preferable as analysts applying the models.

d. A training device should be used in which features could be varied to produce highly effective through to degraded training. The device should be configured to provide three different conditions of training effectiveness (high, mediocre, low); thus, the equivalent of three devices. This would predetermine what each model's predictions should be and would provide an additional criterion to the student ToT scores.

e. Students trained on the device(s) should be at least 15-20 in number per device.

f. Device-trained tasks should be simple but sufficient in number so that a tasks-within-analysts-within-models design would generate enough cases to permit use of regression techniques. Data analysis should determine the relative contribution of each model's component variables (e.g., communality analysis, PC analysis, etc.) to predicting the criterion. This would provide not only more definition of model validity, but would also identify how various aspects of the models function to predict ToT.

g. Reliability assessment should address the summary index, task-level index, and analyst agreement on judgmental (raw) data.

h. Contingent upon outcomes, the study data base should be used to test experimental modifications to constructs and mathematical formulations of the models.

2. No model should be discounted until research as recommended above can fully determine the efficacy of each model and its component variables. The evaluations described in this report have generated numberous empirical questions regarding the models. As these questions are addressed, assumptions supporting the models can be corrected, fine-tuned, and incorporate appropriate new facets or discard existing ones as appropriate to each model's purpose.

3. The efficiency of predictions on device ToT <u>unaided</u> by such models should be determined. In the final analysis, the practical utility of each model will depend upon its ability to enhance the unaided process of predicting superior training device design. The recommended replication study could provide a means for determining unaided and aided ToT prediction accuracy.

REFERENCES

Aagard, J., & Braby, R. (1976). Learning guidelins and algorithms for types of training objectives (Report #23). Orlando, FL: Naval Training Equipment Center, Training Analysis and Evaluation Group.

Branson, R., Raynor, G., Cox, J., & Hannum, W. (1975). Interservice procedures for instructional systems development: Executive summary and model. Ft. Benning, GA: U.S Army Combat Arms Training Board.

Carmines, E. and Zeller, R. (1979). Reliability and Validity Assessment Beverly Hills, CA: Sage Publications, pp. 48-51.

Evans, R., & Swezey, R. (1980). Development of a user's guidebook for TRAINVICE II (Final report for contract # MDA-903-79-C-0428). Alexandria, VA; U.S. Army Research Institute.

Evans, R., & Tufano, D. (1982). Review of four methods for predicting training device effectiveness. Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

Faust, D. (1982). Analysis of learning guidelines used in the prediction of training device effectiveness (Final report, contract # MDA-903-82-M-1878). Alexandria, VA: U.S. Army Research Institute.

Harmon, H. (1967). Modern Factor Analysis. Chicago: University of Chicago Press, pp. 16-19.

Hirshfeld, S., & Kochevar, J. (1979). Training device requirements documents guide. Orlando, FL: PM TRADE, Naval Training Equipment Center.

Kerlinger, F. (1973). Foundations of Behavioral Research. New York, NY: Holt, Rinehart & Winston, pp. 664-65, and 451.

Klein, G., Kane, J., Chinn, A., & Jukes, A. (1978). Analyzing training device effectiveness in cases where test data is inconclusive. Proceedings of the Human Factors Society 22nd Annual Meeting. San Diego, CA: Human Factors Society.

Knerr, C.M., Nadler, L., & Dowell, S.K. (1983). Comparison of training transfer and effectiveness models. Proceedings of the Human Factors Society 27th Annual Meeting. San Diego, CA: Human Factors Society.

Narva, M.A. (1979a). Formative utilization of a model for the prediction of the effectiveness of training devices (Research Memorandum 79-6). Alexandria, VA: U.S. Army Research Institute.

Narva, M.A. (1979b). Development of a systematic methodology for the application of judgmental data to the assessment of training device concepts (Research Memorandum 76-16). Alexandria, VA: U.S. Army Research Institute.

Nunnally, J. (1978). Psychometric Theory. New York, NY: McGraw-Hill.

Rose, A.M., Wheaton, G.R., Leonard, R.L., & Fingerman, P.W. (1976). Evaluation of two tank gunnery trainers (Memorandum). Alexandria, VA: U.S. Army Research Institute.

Shrout, P. and Fleiss, J. (1979) Intraclass correlations: uses in assessing rater reliability. Psychology Bulletin, 86, 420-428.

Swezey, R.W., Chitwood, T.E., Easley, D.L., & Waite, B.J. (1977). Implications for dragon gunnery training development. Springfield, VA: Mellonics Systems Development Division, Litton Industries.

Swezey, R.W., & Evans, R. (1980). Guidebook for users of TRAINVICE II (Contract # MDA-903-79-C-0482). Alexandria, VA: U.S. Army Research Institute.

Swezey, R.W. (1983). Application of a transfer of training model to training device assessment. Journal of Educational Technology Systems, Vol. 11(3).

Thorndike, E.L. (1903). Educational psychology. New York: Lemke and Buechner.

Thorndike and Woodworth (1901). The influence of improvement in one mental function upon the efficiency of other functions. Psychology Review, VII, 247-261, 384-395, 556-564.

Unger, K.W., Swezey, R.W., Hays, R.T., & Mirabella, A. (1984). Army maintenance training and evaluation simulation system (AMTESS) device evaluation, Vol. I, II, III. McLean, VA: Science Applications, Inc.

Wheaton, G.R., Fingerman, P.W., Rose, A.M., & Leonard, R.L., Jr. (1976a). Evaluation of the effectiveness of training devices: Elaboration and application of the predictive model. (Research Memorandum 76-6). Alexandria, VA: U.S. Army Research Institute.

Wheaton, G.R., Rose, A.M., Fingerman, P.W., Korotkin, A.L., & Holding, D. H. (1976b). Evaluation of the effectiveness of training devices: Literature review and preliminary model (Research Memorandum 76-6). Alexandria, VA: U.S. Army Research Institute.

Wheaton, G.R., Rose, A.M., Fingerman, P.W., Leonard, R.L., Jr., and Boycan, G.G. (1976c). Evaluation of three burst-on-target trainers. Alexandria, VA: U.S. Army Research Institute.

Wheaton, G.R., Rose, A.M., Fingerman, P.W., Leonard, R.L., & Boycan, G.G. (1976d). Evaluation of the effectiveness of training devices: Validation of the predictive model. Alexandra, VA: U.S. Army Research Institute.

APPENDIX A

SELECTED CHARACTERISTICS OF THE GRUMMAN APPROACH

Various instructional and technological features of the Grumman simulator are presented in Appendix A. Some of the features which have been incorporated into the missile POI (MOS 24C10) are presented in Table A-1, while Table A-2 lists the lessons which can be taught on the device. A block diagram of the simulator hardware for the 24C10 MOS is presented in Figure A-1.

Instructional and technological features incorporated into the automotive POI (MOS 63D30) are presented in Table A-3. Table A-4 lists the lessons taught in the device. Figures A-2 and A-3 depict the simulator hardware for the 63D30 MOS.

TABLE A-1. AMTESS CAPABILITIES DEMONSTRATED - MISSILE (GRUMMAN)

- TUTORIAL TRAINING VIA VIDEO DISC WITH ADVANTAGE OF STOP ACTION, MOTION, SOUND, VARIATION IN ENRICHMENT OF INSTRUCTIONAL MATERIALS (ADAPTIVE)
- MODELING VIA MOTION
- CUEING/PROMPTING, REMEDIATION
- CAPACITY FOR INSTRUCTIONAL FRAMES 54000 (STILL/MOTION)
- HANDS ON/HEADS ON INTEGRATION OF THE WHOLE TASK (COGNITIVE/MANIPULATIVE ELEMENTS)
- INDIVIDUALLY PACED, PERFORMANCE BASED TRAINING
- DYNAMIC APPLICATION OF TROUBELSHOOTING PROCEDURES (3D COMPONENT)
- INTEGRATION OF JOB PERFORMANCE AIDS
- EFFECTIVE USE OF TRAINING TIME (E.G., DEPENDENCY DIAGRAMS)

TABLE A-2. MISSILE PROGRAM OF INSTRUCTION

LESSON 1: HIGH VOLTAGE CIRCUITS

2: RF GENERATION CIRCUIT

3: ARC DETECTION CIRCUIT

4: OPTION CAPABILITY

— HANDS-ON PRACTICE - XMTR WEEKLY CHECK

— D&P FAULT ISOLATION NOISE DEGENERATION CIRCUIT

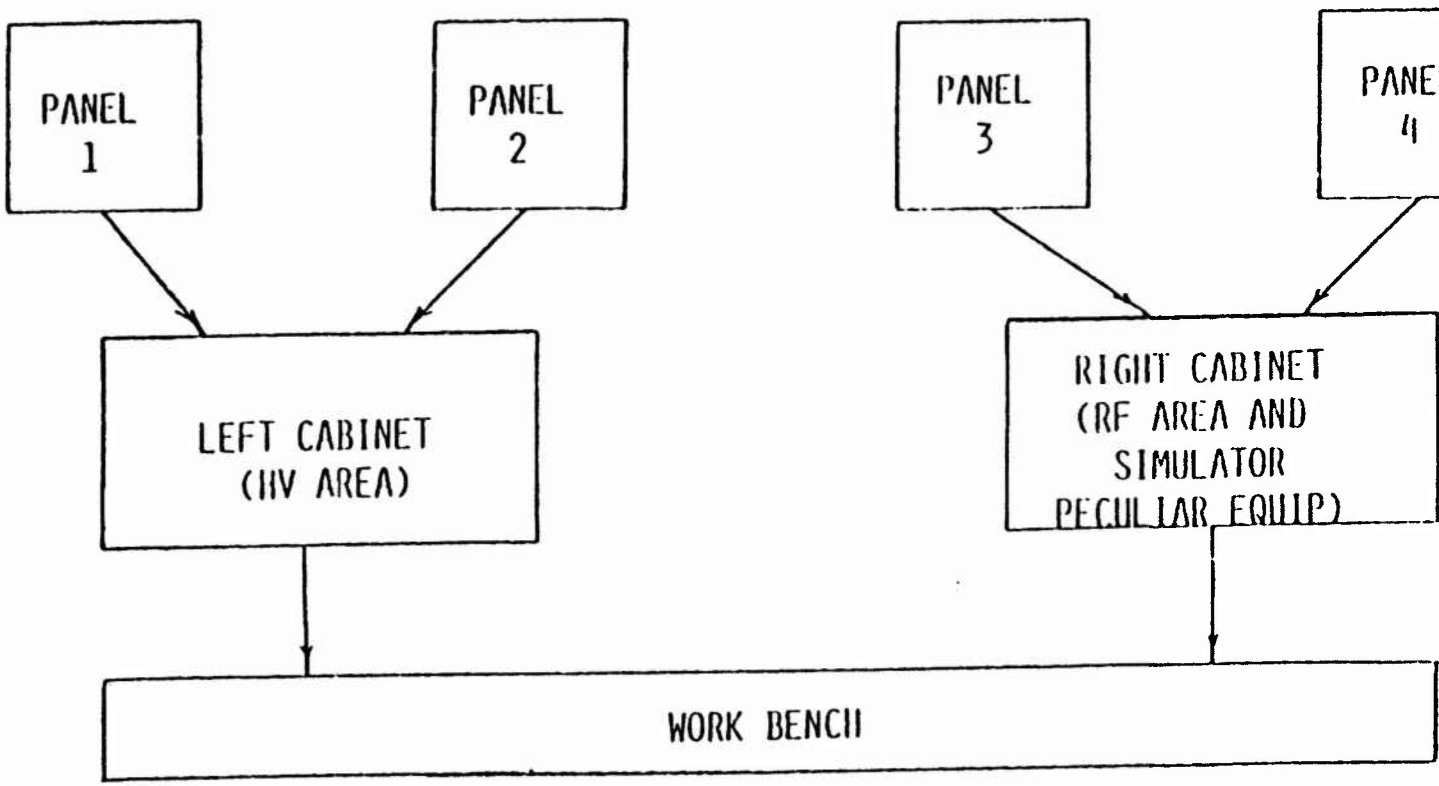


Figure A-1. 3D Missile Mechanical Block Diagram (Grumman)

TABLE A-3. AMTESS CAPABILITIES - AUTOMOTIVE (GRUMMAN)

- GUIDED APPLICATION OF TROUBLESHOOTING
- HIGH FIDELITY AUDITORY CUE PRESENTATION VIA VIDEODISC INTEGRATED WITH HANDS-ON ACTION
- MASTER MODELING OF OPERATIONS IMMEDIATELY FOLLOWED BY APPLICATION
- ADAPTIVE INSTRUCTIONAL MATERIALS
- STUDENT INTERACTION VIA TOUCH PANEL
- HEAVY PICTORIAL/AUDITORY PRESENTATION TO MINIMIZE EFFECT OF ANY READING DEFICIENCIES
- TRAINING OF WHOLE TASK - COGNITIVE/MOTOR ELEMENT — WHAT DONE, WHEN, WHY, AND HOW
- FEEDBACK TO STUDENT

TABLE A-4. AUTOMOTIVE PROGRAM OF INSTRUCTION

LESSON 1: STE/ICE

LESSON 2: TROUBLESHOOT STARTING SYSTEM

LESSON 3: TROUBLESHOOT CHARGING SYSTEM

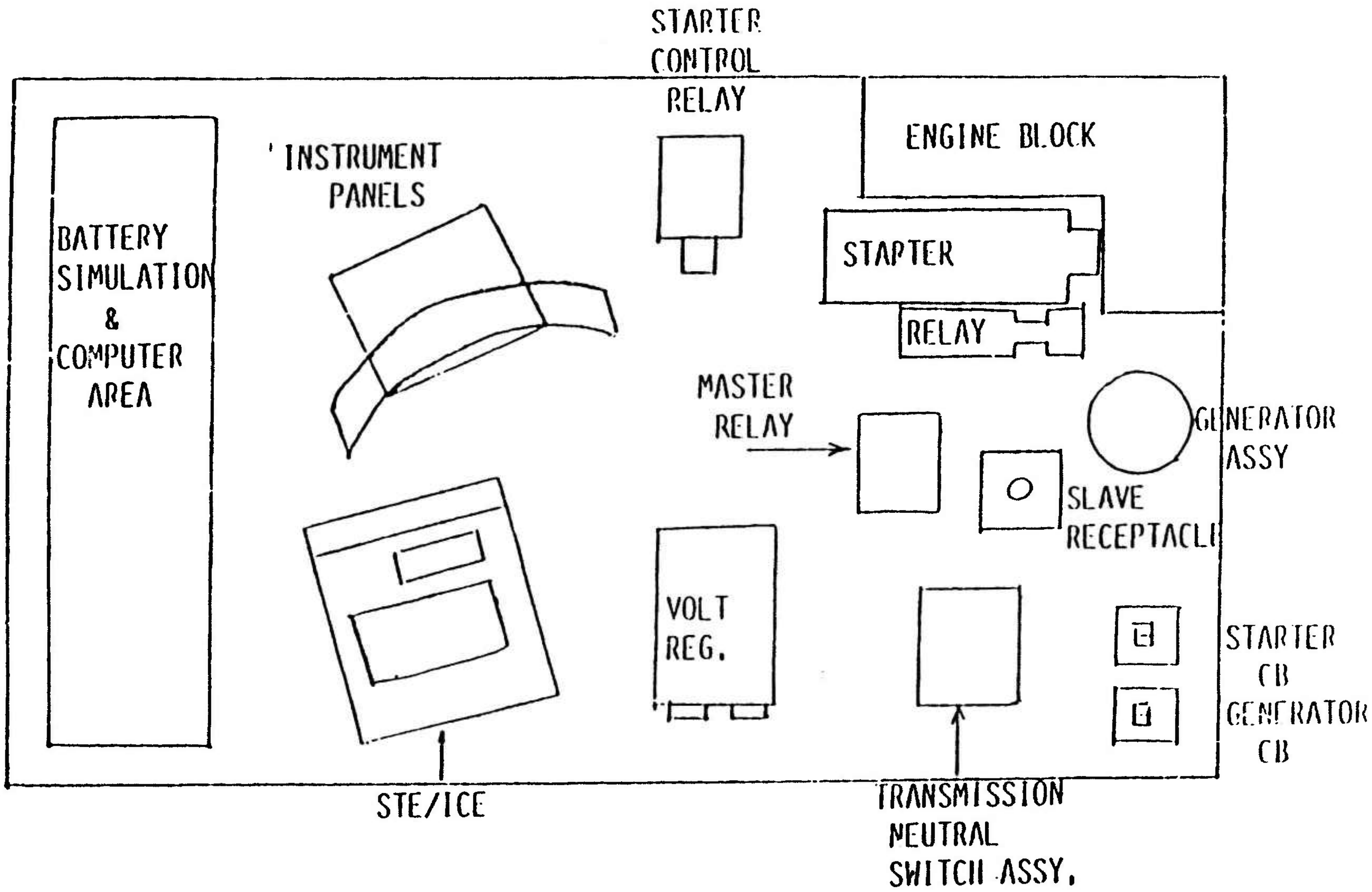


Figure A-2. Automotive 3D Station Layout

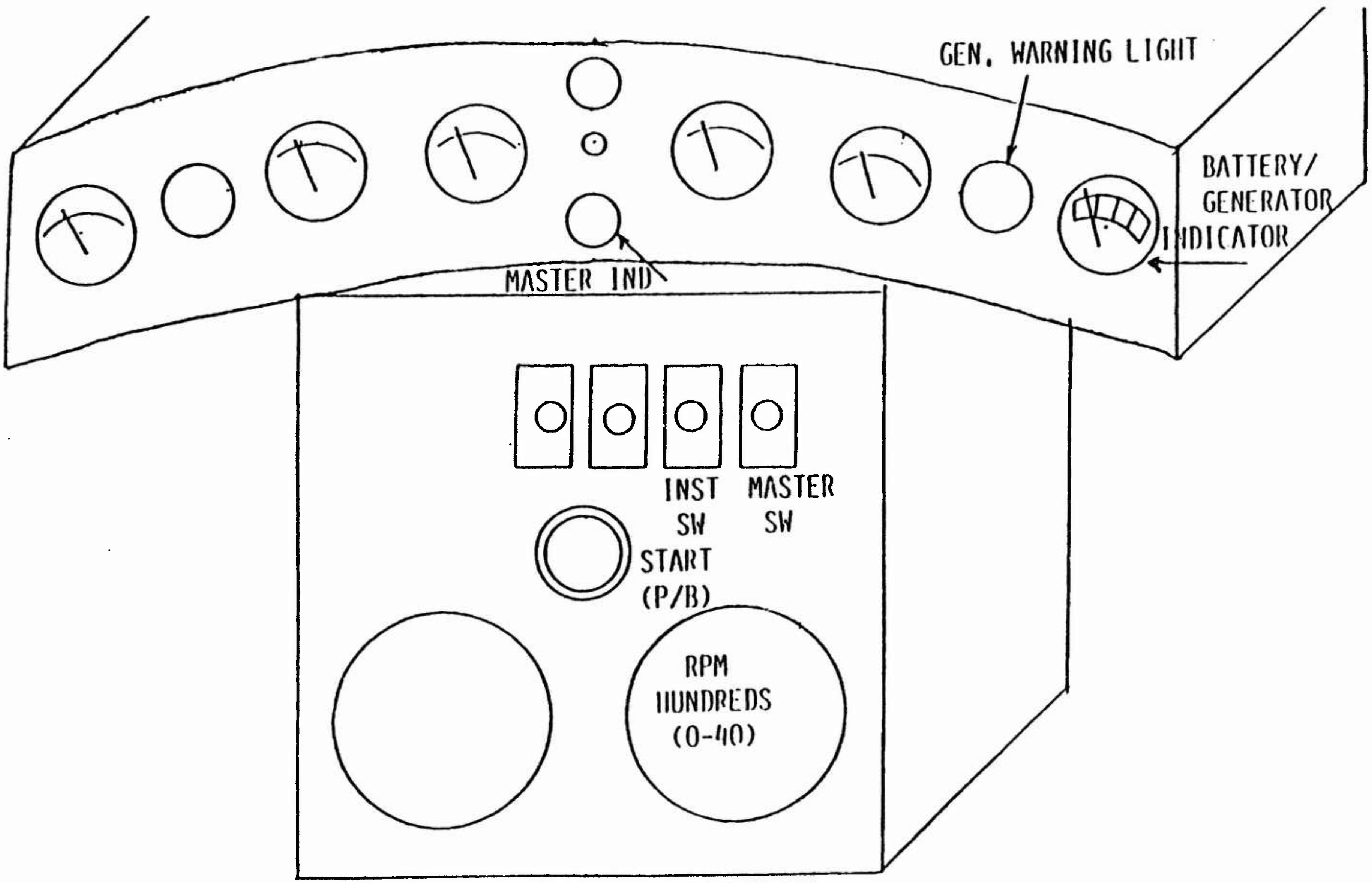


Figure A-3. Grumman Instrument Panel

APPENDIX B

SELECTED CHARACTERISTICS OF THE BURTEK/SEVILLE APPROACH

Simulator hardware for both the 24C10 and the 63W10 MOSs is listed in Table B-1. Figures B-1, B-2, and B-3 illustrate the student station, the instructor control panel, and the student response panel. Table B-2 lists the components of device (for the 63W10 MOS) which students may choose to inspect, remove/replace, or repair/adjust. The basic components of the simulated diesel engine are presented in Figure B-4. Table B-3 lists the exercises which can be taught on the device and Table B-4 lists the various malfunctions which can be induced.

TABLE B-1. BURTEK/SEVILLE AMTESS BREADBOARD MODEL HARDWARE

AUTOMOTIVE SYSTEM	COMMON CORE CONFIGURATION ITEMS:	MISSILE SYSTEM
UNIQUE CONFIGURATION ITEMS:	MASTER CONSOLE	UNIQUE CONFIGURATION ITEMS:
ROLL AROUND SUPPORT STAND	INSTRUCTOR CONTROL PANEL	ROLL AROUND SUPPORT STAND
CYLINDER BLOCK WITH	CRT/KEYBOARD TERMINAL	TRANSMITTER ASSEMBLY
ACCESSORIES	POWER SUPPLIES	BUILT-IN TEST EQUIPMENT
INSTRUMENT PANEL	POWER DISTRIBUTION PANEL	INDICATORS
ENGINE CONTROLS	COMPUTER & I.O.	CONTROLS
REMOVABLE COMPONENTS	WINCHESTER DISK	REMOVABLE COMPONENTS
ADJUSTMENTS & SENSORS		ADJUSTMENTS & SENSORS
SIMULATED TEST EQUIPMENT		SIMULATED TEST EQUIPMENT
STE/ICE	HARD COPY PRINTER	MULTIMETER
		WAVEMETER TEST SET
	TRAINEE CONSOLE	
STANDARD HAND TOOLS	DESK/WORK BENCH	STANDARD HAND TOOLS
(NOT FURNISHED)	CRT DISPLAY UNIT	(NOT FURNISHED)
	PROJECTION SYSTEM	
	TRAINEE RESPONSE PANEL	
	STORAGE DRAWERS	

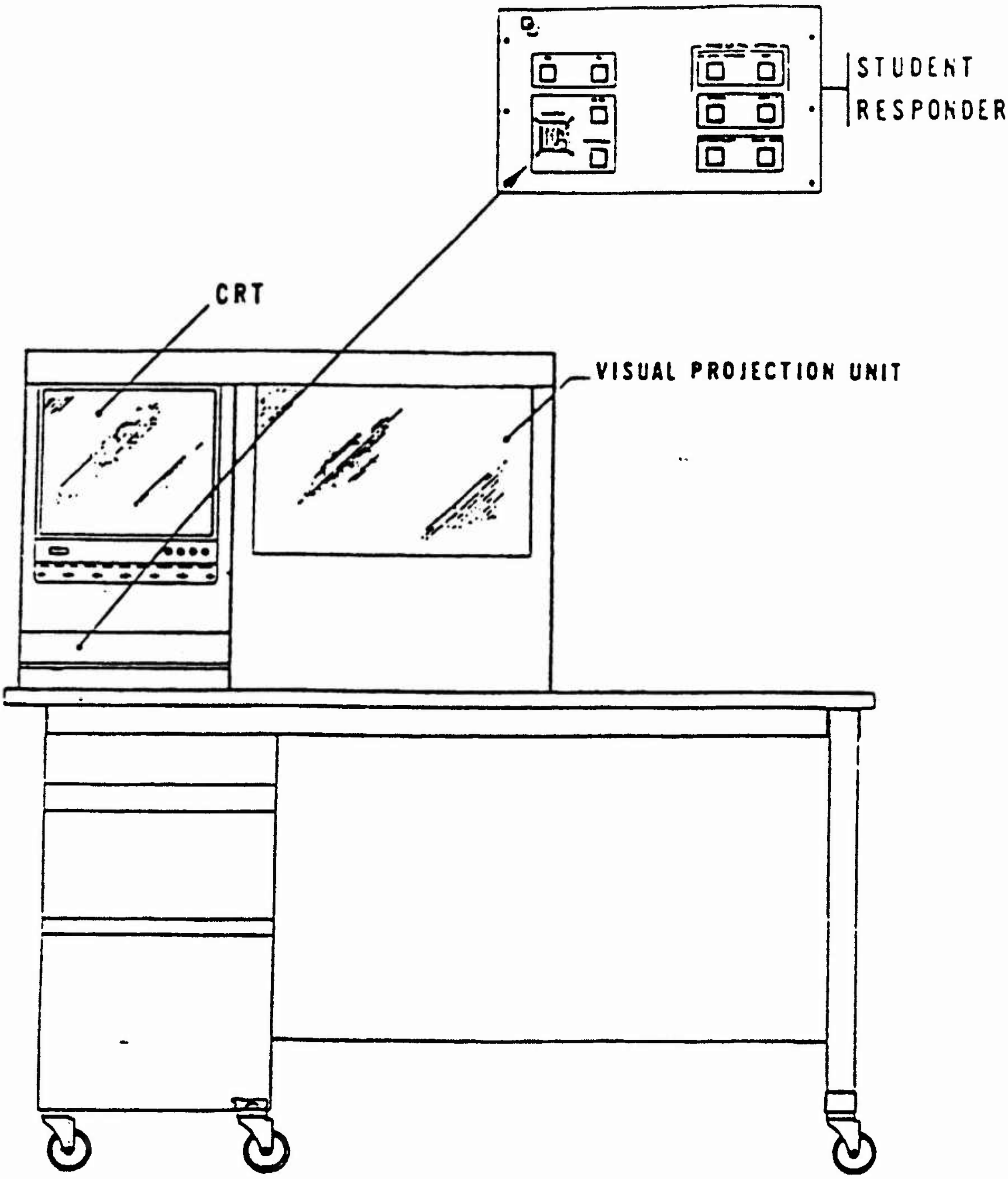


Figure B-1. Student Station of the AMTESS Simulator including CRT Display, Visual Projection Unit, and Student Response Panel (Student Responder)

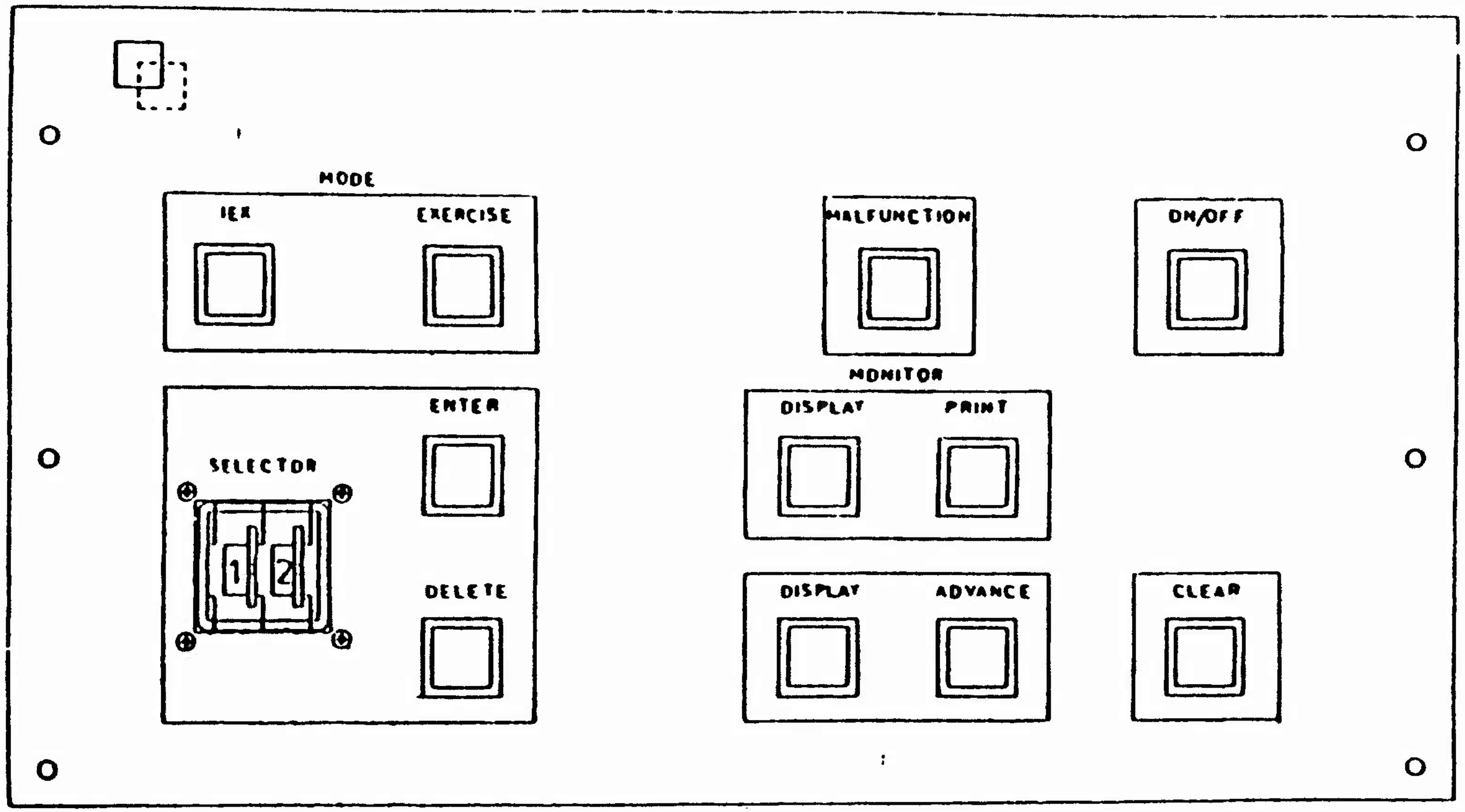


Figure B-2. Controls of the Instructor Control Panel

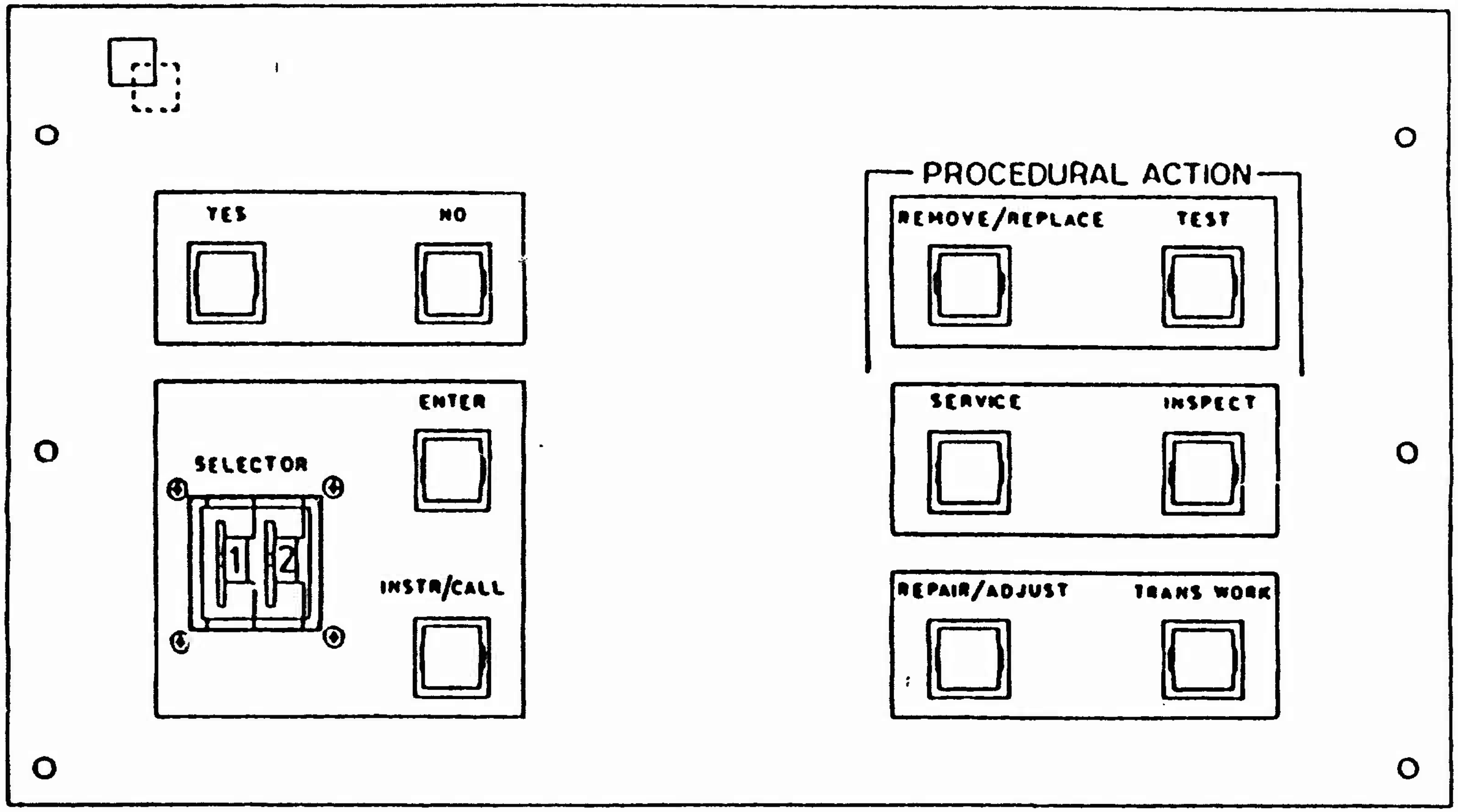
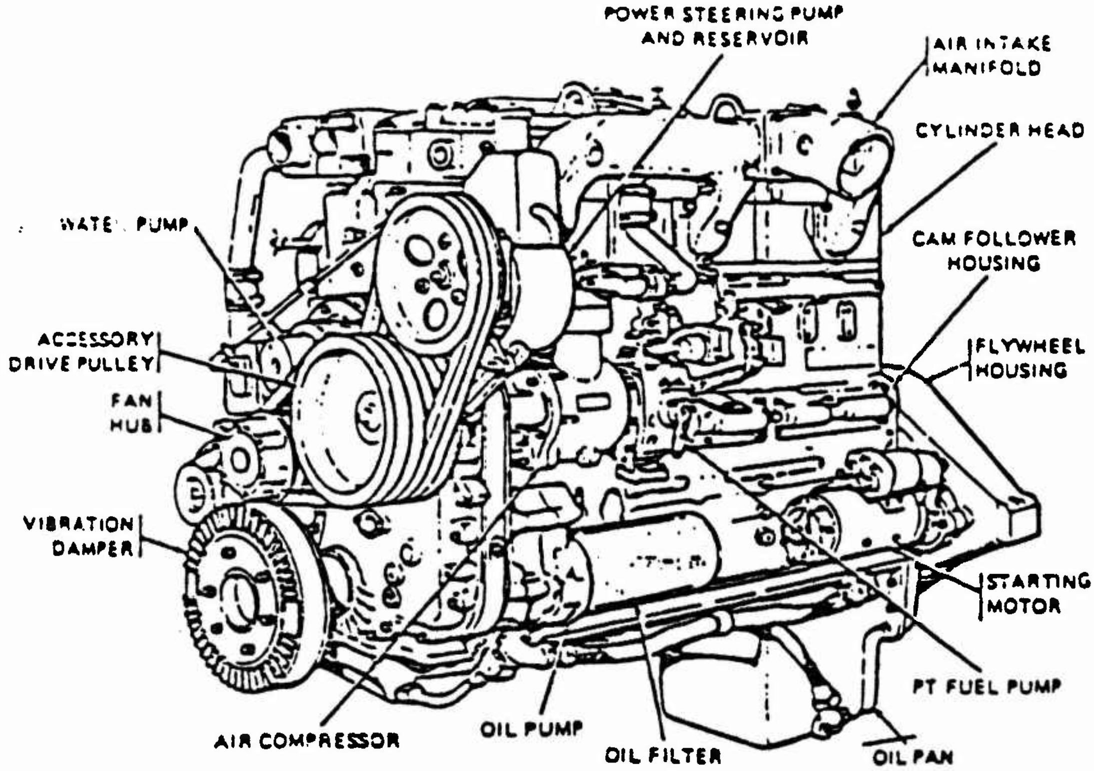


Figure B-3. Controls of the Student Response Panel

TABLE B-2. COMPONENTS OF THE BURTIK/SEVILLE SIMULATOR

01	ALTERNATOR
02	ALTERNATOR DRIVE BELT
03	ALTERNATOR TERMINAL
04	BATTERIES
05	BATTERY CABLES
06	BATTERY CABLE CLAMPS
07	BATTERY ELECTROLYTE
08	BATTERY-GENERATOR INDICATOR
09	BATTERY SWITCH
10	BATTERY TERMINAL CONNECTIONS
11	ELECTROLYTE IMPURITIES
12	ELECTROLYTE SPECIFIC GRAVITY
13	FRONT HARNESS
14	IGNITION SWITCH
15	LEADS AND CONNECTIONS
16	PROTECTIVE CONTROL BOX
17	STARTER AND SOLENOID ASSEMBLY
18	VOLTAGE-ALTERNATOR OUTPUT
19	VOLTAGE-BATTERY SWITCH
20	VOLTAGE-PROTECTIVE CONTROL BOX
21	ELECTRIC FUEL SHUT OFF VALVE
22	FUEL FILTER BODY
23	FUEL FILTER ELEMENT
24	FUEL LINES AND FITTINGS
25	FUEL PUMP
26	PRIMER PUMP
27	PRIMER PUMP NOZZLE
28	OIL
29	OIL FILTER
30	OIL PRESSURE GAGE
31	OIL PRESSURE GAGE PIPING, FITTINGS
32	OIL PRESSURE LOCKOUT SWITCH
33	OIL PUMP
34	OIL PUMP-PICKUP TUBE, RETURN HOSE, MAIN OIL PICKUP HOSE
35	OIL FILTER SHELL
36	COOLANT
37	COOLANT HOSE CLAMPS
38	COOLANT HOSES
39	COOLING SYSTEM
40	FAN DRIVE BELT
41	RADIATOR
42	SURGE TANK
43	THERMOSTAT
44	THERMOSTAT HOUSING GASKET
45	WATER MANIFOLD
46	WATER PUMP
47	WATER PUMP DRIVE BELT



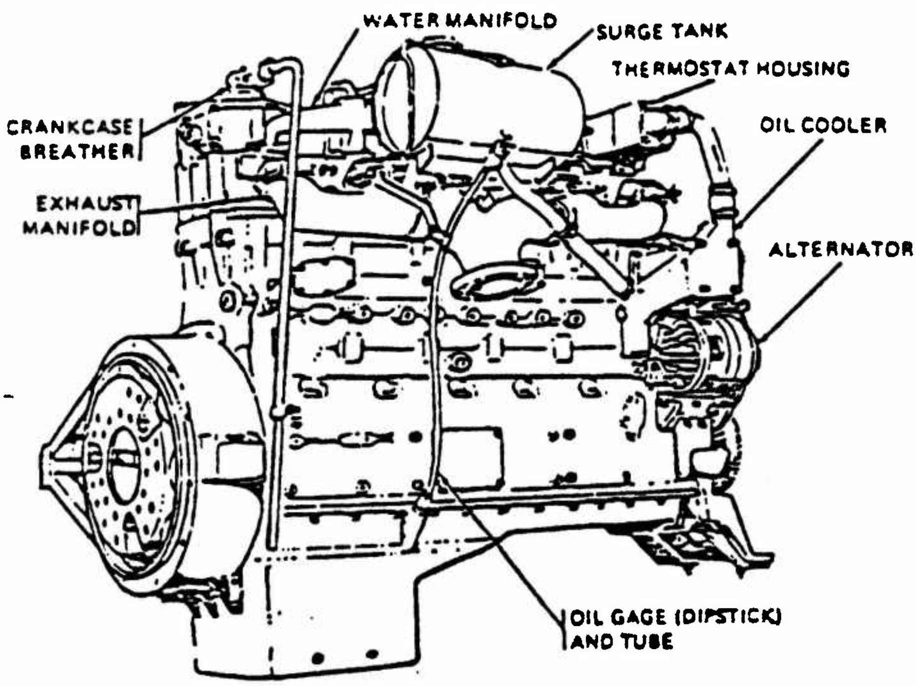


Figure B-4. Basic Components of the NHC-250 Diesel Engine

TABLE B-3. INSTRUCTOR CONTROLLED EXERCISES

01	Normal Operations	
02	Remove & Replace Oil Filter	
03	Remove & Replace Oil Pump	
04	Remove & Replace Thermostat	
05	Remove & Replace Water Pump	
06	Remove & Replace Alternator	
07	Remove & Replace Starter Motor	
08	Remove & Replace Fuel Pump	
09	Adjust Water Pump Drive Belt	
10	Adjust Alternator Drive Belts	
11	Test DC Current	
12	Test Resistance	
13	Test Alternator Output Voltage	
14	Test Oil Pressure	
15	Oil Pump Failure [A]	
16	Oil Pump Failure [C]	
17	Oil Pump Failure [D]	
18	Thermostat Failure [A]	
19	Thermostat Failure [D]	
20	Water Pump Failure [A]	
21	Water Pump Failure [C]	
22	Water Pump Failure [D]	
23	Fuel Pump Failure #1 [A]	no start
24	Fuel Pump Failure #1 [D]	no start

25	Fuel Pump Failure #2 [A]	hard start
26	Fuel Pump Failure #2 [D]	hard start
27	Fuel Pump Failure #3 [A]	eng-stall
28	Fuel Pump Failure #3 [C]	eng-stall
29	Fuel Pump Failure #3 [D]	eng-stall
30	Starter Motor Failure [A]	
31	Starter Motor Failure [C]	
32	Starter Motor Failure [D]	
33	Alternator Failure #1 [A]	hi-charge
34	Alternator Failure #1 [D]	hi-charge
35	Alternator Failure #2 [A]	BG-point-low
36	Alternator Failure #2 [D]	BG-point-low
37	Alternator Failure #3 [A]	low bat
38	Alternator Failure #3 [D]	low bat
39	Alternator Failure #4 [A]	BG-no-move
40	Alternator Failure #4 [D]	BG-no-move
41	Loose Alternator Belt #1 [E]	lo-charge
42	Loose Alternator Belt #2 [E]	BG-point-low
43	Loose Alternator Belt #3 [E]	low bat
44	Loose Alternator Belt #4 [E]	BG-no-move
45	Battery Switch Failure [A]	
46	Battery Switch Failure [D]	
47	Front Harness Failure [A]	
48	Front Harness Failure [D]	
49	Protective Control Box Failure [A]	
50	Protective Control Box Failure [D]	
98	Exercise Continue	
99	Automatic Exercise Reset	

TABLE B-4. MALFUNCTIONS

01	Oil pump
02	Battery Switch
03	Front Harness
04	Protective Control Box
05	Starter Motor
06	Alternator #1
07	Alternator #2
08	Alternator #3
09	Alternator #4
10	Alternator Belt #4
11	Alternator Belt #1
12	Alternator Belt #3
13	Alternator Belt #2
14	Fuel Pump #1
15	Fuel Pump #2
16	Fuel Pump #3
17	Water Pump
18	Thermostat

APPENDIX C
SAMPLE TRAINVICE WORKSHEET
AND RESULTING DATA
(from Swezey and Evans, 1980)

TASKS AND SUBTASKS (with appropriate skills and knowledges)	COVERAGE REQUIREMENTS ANALYSIS	DEVICE I COVERAGE ANALYSIS	DEVICE II COVERAGE ANALYSIS	TRAINING PROFICIENCY ANALYSIS	LEARNING DIFFICULTY ANALYSIS	DEVICE I PHYSICAL CHARACTERISTIC SCORE	DEVICE II PHYSICAL CHARACTERISTIC SCORE	DEVICE I MAXIMUM POSSIBLE PC SCORE	DEVICE II MAXIMUM POSSIBLE PC SCORE	DEVICE I FUNCTIONAL CHARACTERISTIC SCORE	DEVICE II FUNCTIONAL CHARACTERISTIC SCORE	DEVICE I MAXIMUM POSSIBLE FC SCORE	[illegible]

TRAINVICE II MASTER WORKSHEET

I	II	III	IV	V	VI		VII	VIII	IX	X
SKILL/KNOWLEDGE	BEHAVIORAL CATEGORY	LEARNING GUIDELINES FOR GOOD INSTRUCTIONAL PRACTICE	DISPLAY/CONTROL	APPLICABLE GENERIC CHARACTERISTICS	GENERAL CHARACTERISTICS	RATING	PHYSICAL CHARACTERISTICS SCORE SUM OF RATINGS / PC_{max}	APPLICABLE FUNCTIONAL GUIDELINES	RATINGS	FUNCTIONAL CHARACTERISTICS SCORE OVER FC_{max}

TRAINVICE II:

TASKS AND SUBTASKS (with appropriate skills and knowledges)	COVERAGE REQUIREMENTS ANALYSIS	SHERMAN COVERAGE ANALYSIS	PATTON COVERAGE ANALYSIS	TRAINING PROFICIENCY ANALYSIS	LEARNING DIFFICULTY ANALYSIS	SHERMAN PHYSICAL CHARACTERISTIC SCORE	PATTON PHYSICAL CHARACTERISTIC SCORE	SHERMAN MAXIMUM POSSIBLE PC SCORE	PATTON MAXIMUM POSSIBLE PC SCORE	SHERMAN FUNCTIONAL CHARACTERISTIC SCORE	PATTON FUNCTIONAL CHARACTERISTIC SCORE	SHERMAN MAXIMUM POSSIBLE FC SCORE	PATTON MAXIMUM POSSIBLE FC SCORE
1. Load the Launch Tube	CR	C	C	P	D	PC	PC	PC_{max}	PC_{max}	FC	FC	FC_{max}	FC_{max}
1.1 Lock traversing unit in azimuth and elevation													
1.1.1 Operate traversing unit	1	1	1	3	3	10	10	10	10	29	13	10	15
1.1.2 Azimuth and elevation movement	1	1	1	3	3	15	15	15	15	16	15	15	15
1.1.3 Locking mechanism	1	1	1	2	2	15	15	15	15	15	10	10	15
1.2 Remove encased missile from stowed position													
1.2.1 Remove casing materials from missile	0	0	0	0	0	0	0	0	0	0	0	0	0
1.2.2 Missile configuration	0	0	0	0	0	0	0	0	0	0	0	0	0
1.2.3 Casing materials	0	0	0	0	0	0	0	0	0	0	0	0	0
1.3 Load encased missile													
1.3.1 Loading procedure	1	1	1	4	2	10	10	10	10	15	15	15	15
1.3.2 Locking procedure	1	1	1	0	2	10	10	10	10	15	15	15	15
1.3.3 Safety aspects of launch tube preparation	1	1	1	4	2	10	10	10	10	15	15	15	15
1.3.4 Loading and locking mechanisms	1	1	1	0	2	10	10	10	10	15	15	15	15
2. Select a Target													
2.1 Visually select target													
2.1.1 Discrimination of enemy-targets from other targets	0	0	0	0	0	0	0	0	0	0	0	0	0
2.1.2 Silhouettes of threat vehicles	0	0	0	0	0	0	0	0	0	0	0	0	0
2.1.3 Scanning techniques	1	1	1	4	2	13	21	21	21	12	10	10	10
2.2 Swing traversing unit to align optical sight													
2.2.1 Unlocking procedures	1	1	1	4	3	10	10	10	10	15	15	10	10
3. Connect Encased Missile													
5.1 Lift trigger protective cover													
5.1.1 Unlock trigger protective cover	0	0	1	0	0	0	0	0	0	0	0	0	0
5.1.2 Trigger protective cover location	0	0	1	0	0	0	0	0	0	0	0	0	0
5.1.3 Unlocking procedures	0	0	0	0	0	0	0	0	0	0	0	0	0
5.2 Press firing trigger													
5.2.1 Perform firing procedures	1	1	1	4	3	21	10	21	21	24	10	24	21
5.2.2 Firing trigger location	1	1	1	4	3	21	10	21	21	24	10	24	24
5.2.3 Missile firing procedures	1	1	1	4	3	21	15	21	21	24	15	24	24
6. Track Target Until Missile Impact													
6.1 Make continuous adjustments to keep crosshairs centered on target													

FIGURE II-1: COMPLETED TRAINVICE II MASTER WORKSHEET

APPENDIX D
REVISED CRITERION TEST
FOR MOS 63W10

MOS 63 W 10 (WHEELED VEHICLE MECHANIC) PERFORMANCE MEASURES

I. BACKGROUND DATA

STUDENT NAME: ______________________ CLASS # ______ GROUP # ______

GRADE: (E-1, E-2, Other) ______

INSTRUCTOR (CLASSROOM) ______________ TESTING ______

EXP. CONDITION: CONVENTIONAL _____ EXPERIMENTAL _____

DATE: ____/____/82 TIME STARTED ______________

ATTEMPT # 1____ 2____ 3____ GRADE: PASS ____ FAIL ____

INDUCED MALFUNCTION ____OIL PUMP FAILURE____

GO	NO GO		COMMENTS
		TROUBLESHOOTING ENGINE MALFUNCTION TIME STARTED ______________ 1. Determine malfunction. TIME FINISHED ______________	
		TIME STARTED ______________	
		1. Select TM 9-2320-260-20-2-1, pg. 6-2.	
		2. Select low or no oil pressure, pg. 8-2. Check oil pressure gauge piping and fitting.	
		3. Signs of leaking oil.	
		4. Bent, cracked or broken piping.	
		5. Loose fittings. Check service ability of oil pressure gauge (describe to instructor using TM AS NEEDED). NOTE CAUTION.	
		6. Remove oil pressure pipe.	
		7. Screw on test gauge. 16 PSI	
		8. Start engine.	
		9. Refer to 260-10-2, pp. 1-16 (15 to 20 PSI).	
		10. See if test gauge pressure is higher. (If reading stays low, tell direct support)	

TIME FINISHED ______________

GO NO GO TIME STARTED ______________

1. Select TM 9-2320-260-34-1.
2. Select low or no oil pressure, pg. 3-2.
3. Check for loose fittings.
4. Check for leaking hoses.
5. Check for broken pickup tube.
6. Are the above three in functioning order?
7. Correctly use manual to determine need for oil pump removal.

TIME FINISHED ______________

OIL PUMP FILTER AND PUMP REMOVAL

Removal

TIME STARTED ______________

1. Select TM 9-2320-34-2-1, pg. 3-182.
2. Select TM 9-2320-34-21, pg. 2-29 or TM 9-2320-260-20.
3. Remove center bolt. 9/16" wrench
4. Remove filter assembly.
5. Indicate throw away filter element and seal.

TIME FINISHED ______________

Oil pump

TIME STARTED ______________

1. Select pg. 3-183.
2. Remove two bolts, washers, and hose clamps. 5/8" wrench.
3. Remove return hose. 1-1/4" and 1-1/2" wrenches.
4. Remove elbow tube. 1-1/4" wrench.
5. Remove pickup hose. 1-1/4" and 1-3/8" wrenches.
6. Remove four bolts and lockwashers. 5/8" wrench.
7. Remove one bolt (centerline). 5/8" wrench.
8. Remove oil pump and gasket.
9. Indicate throw away gasket.

TIME FINISHED ______________

SUBJECT #______

MOS 63 W 10

STUDENT NAME: ______________

GO	NO GO		COMMENTS
		OIL PUMP FILTER AND PUMP REPLACEMENT	
		Replacement	
		TIME STARTED ____________	
		1. Select TM 9-2320-260-34-2-1, pg. 3-196.	
		2. Place gasket on pump body.	
		3. Screw in and tighten two bolts and lockwashers to pump plate. 5/8" wrench.	
		4. Screw in and tighten bolt and lockwasher (center line), 5/8" wrench.	
		5. Screw in and tighten 6-1/2" bolt (very top) and lockwasher. 5/8" wrench.	
		6. Screw in and tighten 7-1/2" bolt (bottom, behind filter) and lockwasher. 5/8" wrench.	
		7. Replace pickup (short) hose. 1-1/4" and 1-3/8" wrenches	
		8. Replace elbo tube. 1-1/4" wrench	
		9. Replace return (long) hose. 1-1/4" wrench.	
		10. Replace two clamps, bolts, washers. 5/8" wrench.	
		11. Replace seal and filter element and assembly.	
		12. Replace center bolt. 9/16" wrench	
		TIME FINISHED ____________	

APPENDIX E
ANALYST OPINION QUESTIONNAIRE

TRAINVICE QUESTIONNAIRE

Name ________________

We are interested in determining how you feel about the TRAINVICE models in terms of difficulty and effectiveness. Please complete the following ratings for each of the analyses involved in the TRAINVICE models according to the following five-point scales.

DIFFICULTY

RATING	DEFINITION
5	Ratings were very difficult to make
4	Ratings were somewhat difficult to make
3	Ratings were neither difficult nor easy
2	Ratings were somewhat easy to make
1	Ratings were easy to make

EFFECTIVENESS

RATING	DEFINITION
5	Ratings were very effective in assessing the simulator
4	Ratings were somewhat effective in assessing the simulator
3	Unsure of the effectiveness of the ratings
2	Ratings were somewhat ineffective in assessing the simulator
1	Ratings were highly ineffective in assessing the simulator

TRAINVICE I

ANALYSIS	DIFFICULTY	EFFECTIVENESS	COMMENTS
Task Commonality	______	______	______
Physical Similarity	______	______	______
Functional Similarity	______	______	______
Learning Deficit	______	______	______
Training Techniques	______	______	______

TRAINVICE II

ANALYSIS	DIFFICULTY	EFFECTIVENESS	COMMENTS
Task Commonality	______	______	______
Physical Similarity	______	______	______
Functional Similarity	______	______	______
Skill & Knowledge Requirements	______	______	______
Task Training Difficulty	______	______	______

TRAINVICE QUESTIONNAIRE

Name ____________

TRAINVICE III & IV

ANALYSIS	DIFFICULTY	EFFECTIVENESS	COMMENTS
Coverage Requirement			
Coverage			
Training Proficiency			
Learning Difficulty			
Physical Characteristics			
Functional Characteristics			

APPENDIX F
MOS TASKS/SUBTASKS
INVOLVED IN THE STUDY

MOS 63D30/H30 TASKS

TASK #1: START ENGINE AND CONFIRM GENERATOR WARNING LIGHT ON

SUBTASKS:

1. Set vehicle parking brake.
2. Transmission lever in neutral and locked.
3. Push throttle control in.
4. Set master switch on.
5. Set instrument switch on.
6. Check master indicator light on.
7. Push in start switch and hold until engine starts.
8. Indicator generator warning light on.
9. Check generator indicator gauge in the green.
10. Pull out engine shutdown handle with engine stops.
11. Set instrument switch off.
12. Set master switch off.

TASK #2: PERFORM VTM HOOK-UP AND CHECK-OUT

SUBTASKS:

1. Pull off power switch on the VTM.
2. Connect P1 of the power cable W5 to J1 on the VTM.
3. Connect the red clip lead of cable W5 to the positive terminal of vehicle battery.
4. Connect black clip lead of cable W5 to the negative terminal of vehicle battery.
5. Push on the power switch on the VTM.
6. Verify that display indicates .8.8.8.8 for 2 seconds then changes.
7. Dial 66 into test select and press test.
8. Verify that VTM displays and holds "0066."
9. Dial test select to 99 and press test.
10. Verify that VTM displays 099, blank, .8.8.8.8, blank, several numbers then displays and holds "Pass."
11. Dial 60 into test select and press test.
12. When "VEH" appears, dial "10" into test select.
13. Press test switch and ensure VTM displays "10."

MOS 63W10 TASKS

TASK #1: TROUBLESHOOT ENGINE MALFUNCTION

SUBTASKS:

1. Start engine.
2. Stop engine.
3. Check oil pipes for leaking oil.
4. Check oil pipes for bends, cracks, brakes.
5. Check oil pressure gauge piping and...
6. Remove oil pressure pipe.
7. Screw on test gauge.
8. Start engine.
9. Determine if test gauge pressure...
10. Stop engine.
11. If reading stays low, tell direct...

TASK #2: OIL PUMP FILTER AND PUMP REPLACE

SUBTASKS:

1. Place gasket on pump body.
2. Place pump onto engine.
3. Screw in and tighten 2 bolts and...
4. Screw in and tighten bolt and...
5. Screw in and tighten 6-1/2" bolt...
6. Screw in and tighten 7-1/2" bolt...
7. Replace pickup hose.
8. Replace elbo tube.
9. Replace return hose.
10. Replace 2 clamps, bolts, washers.
11. Replace seal and filter element.
12. Replace center bolt.

MOS 24C10 TASKS

TASK #1: CHECK ION TEST

SUBTASKS:

1. Verify that BATTLE SHORT switch is set to NORMAL.
2. Perform the interlock bypass procedure.
3. Set ION PROBE TEST switch to POS 2, then release switch.
4. Press and release RADIATE pushbutton.
5. Press and release RADIATE INTLK RESET pushbutton.
6. Press and release RADIATE pushbutton.
7. Close and secure Transmitter Panel 3.

TASK #2: CHECK MASTER OSCILLATOR AND POWER AMPLIFIER

SUBTASKS:

1. Press and release STANDBY pushbutton.
2. Set Master oscillator BEAM circuit breaker to ON.
3. Set power amplifier BEAM circuit breaker to ON.
4. Set REGULATOR VOLTS switch to MO.
5. Press and release radiate pushbutton.
6. Set REGULATOR VOLTS switch to PA.
7. Set REGULATOR VOLTS switch to OFF.
8. Set transmitter test set SELECTOR switch (11, fig. 2-8) to position 2 (XTAL BALANCE).
9. Set Forward rf power switch to PA.
10. Press and hold ARC DETECTOR TEST pushbutton.
11. ARC DETECTOR TEST pushbutton release.
12. Observe REFLECTED RF POWER meter is in green area.

TASK #3: CHECK LOCAL OSCILLATOR CRYSTAL CURRENT

SUBTASKS:

1. Set Degeneration function Selector switch to Lo Power.
2. Observe degeneration function monitor meter is steady in the upper orange area.

TASK #4: CHECK REFERENCE LEVEL

SUBTASKS:

1. Set Degeneration function selector switch to REF Level.
2. Observe Degeneration function monitor meter indication remains stable in orange or green area.

APPENDIX G

TASK-LEVEL PREDICTIONS AND PAIRED ToT CRITERION MEASURES

NOTE: For each model, X = task-level prediction metric, Y = ToT mean for the task (criterion)

DEVICES, MOS	TRAINING TASK	ANALYST (Ss)	CASFS	(Wheaton, et al.) X	(Wheaton, et al.) Y	(Hirshfeld, Kochevar) X	(Hirshfeld, Kochevar) Y	(Narva) X	(Narva) Y	(Swezey, Evans) X	(Swezey, Evans) Y
• Grumman (63D/H)	#1	S_1	1	38	89	84	89	32	89	78	89
		S_2	2	8	89	80	89	28	89	75	89
		S_3	3	50	89	81	89	31	89	87	89
	#2	S_1	4	46	94	98	94	25	94	83	94
		S_2	5	32	94	95	94	28	94	87	94
		S_3	6	0	94	98	94	13	94	82	94
• Seville/ Burtek (63W)	#1	S_1	7	12	45	68	45	18	45	80	45
		S_2	8	15	45	68	45	7	45	57	45
		S_4	9	21	45	72	45	41	45	77	45
		S_5	10	25	45	66	45	20	45	80	45
	#2	S_1	11	27	91	95	91	20	91	82	91
		S_2	12	47	91	85	91	15	91	56	91
		S_4	13	40	91	96	91	46	91	81	91
		S_5	14	62	91	91	91	17	91	70	91
• Grumman (24C)	#1	S_6	15	33	90.3	93	90.3	29	90.3	81	90.3
		S_7	16	36	90.3	63	90.3	5	90.3	81	90.3
	#2	S_6	17	45	100	98	100	38	100	75	100
		S_7	18	47	100	63	100	5	100	77	100
	#3	S_6	19	55	91	100	91	56	91	70	91
		S_7	20	54	91	63	91	5	91	73	91
	#4	S_6	21	55	90	100	90	12	90	81	90
		S_7	22	54	90	63	90	5	90	81	90
• Seville/ Burtek (24C)	#1	S_6	23	19	93	36	93	31	93	85	93
		S_8	24	19	93	29	93	35	93	87	93
		S_9	25	35	93	17	93	25	93	90	93
	#2	S_6	26	25	98	58	98	42	98	83	98
		S_8	27	29	98	33	98	24	98	85	98
		S_9	28	24	98	35	98	33	98	91	98
	#3	S_6	29	29	84	63	84	47	84	83	84
		S_8	30	28	84	53	84	21	84	85	84
		S_9	31	25	84	50	84	37	84	93	84
	#4	S_6	32	29	100	63	100	23	100	93	100
		S_8	33	28	100	54	100	23	100	93	100
		S_9	34	25	100	50	100	20	100	82	100